1761

[illegible]

1.7541 Seconds

[illegible] updates/sec

[illegible]

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	[illegible]	[illegible]	1	AY[illegible]	AY[illegible] Borrelia
2	21	[illegible]	[illegible]	1	[illegible]	[illegible] Borrelia
3	21	[illegible]	[illegible]	1	A[illegible]	A[illegible] Borrelia
4	21	100.0	29158	9	AC[illegible]	A[illegible] Homo sap.
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]	A[illegible]	A[illegible]
7	[illegible]	[illegible]	[illegible]	2	A[illegible]	A[illegible]
8	19.4	92.4	54245	2	A[illegible]	A[illegible] Mus mus
9	19.4	92.4	[illegible]	[illegible]	A[illegible]	A[illegible] Homo sap.
10	19	[illegible]	[illegible]	2	A[illegible]	A[illegible]
11	19	[illegible]	[illegible]	2	A[illegible]	A[illegible]
12	18.4	87.6	[illegible]	[illegible]	[illegible]	[illegible]
13	18.4	87.6	[illegible]	2	A[illegible]44	A[illegible]44 Homo sap.
14	18.4	87.6	[illegible]	[illegible]	A[illegible]	A[illegible] Homo sap.
15	18.4	87.6	[illegible]	[illegible]	A[illegible]	A[illegible]
16	18.4	87.6	[illegible]	2	A[illegible]44	A[illegible]44
17	18.4	87.6	[illegible]	2	A[illegible]	A[illegible]
18	18.4	87.6	[illegible]	2	A[illegible]	A[illegible]
19	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Homo sap.
20	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Homo sap.
21	18.4	87.6	[illegible]	2	A[illegible]	A[illegible]
22	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Mus
23	18.4	87.6	[illegible]	2	A[illegible]	A[illegible]
24	18.4	87.6	[illegible]	2	A[illegible]	A[illegible]
25	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Homo sap.
26	18.4	87.6	[illegible]	[illegible]	A[illegible]	A[illegible] Homo sap.
27	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Homo sap.
28	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Homo sap.
29	18.4	87.6	[illegible]	2	A[illegible]444	A[illegible]444 Mus mus
30	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Homo sap.
31	18.4	87.6	[illegible]	[illegible]	A[illegible]	A[illegible] Homo sap.
32	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Mus mus
33	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Mus mus
34	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Mus mus
35	18.4	87.6	[illegible]	2	A[illegible]	A[illegible] Homo sap.
36	18	[illegible]	[illegible]	[illegible]	AL[illegible]	AL[illegible] Mus mus
37	18	[illegible]	[illegible]	2	A[illegible]	A[illegible]
38	18	[illegible]	[illegible]	9	HS[illegible]	[illegible] Human
39	18	[illegible]	[illegible]	9	A[illegible]	A[illegible] Homo sap.
40	18	[illegible]	[illegible]	[illegible]	A[illegible]	A[illegible]
41	18	[illegible]	[illegible]	2	A[illegible]	A[illegible] Mus mus
42	18	[illegible]	172[illegible]	2	A[illegible]	A[illegible]
43	18	[illegible]	174[illegible]	[illegible]	A[illegible]	A[illegible]
44	18	[illegible]	[illegible]	[illegible]	[illegible]	A[illegible] Human
45	18	[illegible]	[illegible]	2	A[illegible]	A[illegible]

ALIGNMENTS

RESULT 1
AY[illegible]
LOCUS AY[illegible] DNA [illegible]
DEFINITION Borrelia burgdorferi [illegible] A (lipA), [illegible] complete cds.
ACCESSION AY[illegible]
VERSION AY[illegible]
KEYWORDS .
SOURCE Borrelia burgdorferi.
ORGANISM Borrelia burgdorferi
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.
REFERENCE 1 (bases 1 to [illegible])
AUTHORS [illegible], Sixen,H., [illegible] and [illegible]
TITLE [illegible]

```
            borreliosis
  JOURNAL   J. Clin. Microbiol. 39 (11), 4013-4019 (2001)
  MEDLINE   21529965
   PUBMED   11682523
REFERENCE   2  (bases 1 to 1026)
  AUTHORS   Panelius,J., Lahdenne,P., Saxen,H., Heikkila,T. and Seppala,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (26 SEP 2000) Haartman Institute, University of Helsinki
            Haartmaninkatu 3, Helsinki 00290, Finland
FEATURES             Location/Qualifiers
     source          1..1026
                     /organism="Borrelia burgdorferi"
                     /strain="IA"
                     /db_xref="taxon:139"
                     /note="sensu stricto"
     gene            1..1026
                     /gene="flaA"
     CDS             1..1026
                     /gene="flaA"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellin A"
                     /protein_id="AAL25582.1"
                     /db_xref="GI:[illegible]"
                     /translation="[illegible]FAQ[illegible]
                     AE[illegible]MVVI[illegible]INNW[illegible]QATVEN[illegible]V
                     VAPAVVE[illegible]SAM[illegible]
                     [illegible]NIKIMEI[illegible]MEYA[illegible]NYIP
                     N[illegible]HS[illegible]FLSVSID
                     [illegible]VL[illegible]NEVLK[illegible]
                     EESSPKN"
BASE COUNT      329 a    137 c    207 g    353 t
ORIGIN

  Query Match             100.0%;  Score 21;  DB 1;  Length 1026;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAAAGT 21
              |||||||||||||||||||||
Db          1 ATGAAAAGGAAAGCTAAAAGT 21


RESULT 2
BB0629[illegible]
LOCUS       BBU62900                1655 bp    DNA     linear   BCT 15-JAN-1997
DEFINITION  Borrelia burgdorferi flagellar filament outsheath protein (flaA)
            gene, complete cds, and chemotaxis histidine kinase (cheA) gene,
            partial cds.
ACCESSION   U62900
VERSION     U62900.1  GI:1575445
KEYWORDS    .
SOURCE      Borrelia burgdorferi.
  ORGANISM  Borrelia burgdorferi
            Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE   1  (bases 1 to 1655)
  AUTHORS   Ge,Y. and Charon,N.W.
  TITLE     An unexpected flaA homolog is present and expressed in Borrelia
            burgdorferi
  JOURNAL   J. Bacteriol. 179 (2), 552-556 (1997)
  MEDLINE   97144545
   PUBMED   8990312
REFERENCE   2  (bases 1 to 1655)
  AUTHORS   Ge,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1996) Yigong Ge, West Virginia University,
            Microbiology, HSCN, Morgantown, WV 26506, USA
FEATURES             Location/Qualifiers
     source          1..1655
                     /organism="Borrelia burgdorferi"
                     /strain="212"
                     /db_xref="taxon:139"
                     /clone="pW1 and pW2"
     CDS             <1..395
                     /function="unknown"
                     /note="orfA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAC44769.1"
                     /db_xref="GI:1575446"
                     /translation="[illegible]
                     [illegible]
                     IKIMRENYP[illegible]ILVER[illegible]ILKEQS"
     gene            473..1498
                     /gene="flaA"
     CDS             473..1498
                     /gene="flaA"
                     /function="putative flagellar filament outsheath protein"
                     /codon_start=1
                     /transl_table=11
                     /product="FlaA protein"
                     /protein_id="AAC44770.1"
                     /db_xref="GI:1575447"
                     /translation="MKKKAK[illegible]
                     AE[illegible]NNW[illegible]QATVENSV
                     VAPAVVE[illegible]
                     [illegible]NYIP
                     NI[illegible]FLSVSID
                     [illegible]NEVLK[illegible]
                     EESSPKN"
     gene            1566..>1655
                     /gene="cheA"
     CDS             1566..>1655
                     /gene="cheA"
                     /function="chemotaxis histidine kinase"
                     /codon_start=1
                     /transl_table=11
                     /product="CheA protein"
                     /protein_id="AAC44771.1"
                     /db_xref="GI:1575448"
                     /translation="MEIIDLENEEI[illegible]QNLVE[illegible]"
BASE COUNT      560 a    182 c    308 g    605 t
ORIGIN

  Query Match             100.0%;  Score 21;  DB 1;  Length 1655;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAAAGT 21
              |||||||||||||||||||||
Db        473 ATGAAAAGGAAAGCTAAAAGT 493


RESULT 3
AE001168
LOCUS       AE001168               13271 bp    DNA     linear   BCT 17-DEC-1997
DEFINITION  Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION   AE001168 AE000783
VERSION     AE001168.1  GI:2688598
KEYWORDS    .
SOURCE      Borrelia burgdorferi.
  ORGANISM  Borrelia burgdorferi
            Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE   1  (bases 1 to 13271)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.,
            Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
            Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D.,
            Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S.,
            Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.,
            Venter,J.C. et al.
  TITLE     Genomic sequence of a Lyme disease spirochaete, Borrelia
            burgdorferi
  JOURNAL   Nature 390 (6660), 580-586 (1997)
  MEDLINE   98065943
```

[illegible]

identified by sequence similarity; putative"

[illegible]

/db_xref="[illegible]"
/translation="[illegible]
FIKEKSIKYFNIYEKV"
gene [illegible]
/gene="[illegible]"
CDS [illegible]
/gene="[illegible]"
/note="[illegible]
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="[illegible]"
/protein_id="AA[illegible]"
/db_xref="[illegible]"
/translation="[illegible]
V"
gene [illegible]
/gene="[illegible]"
CDS [illegible]
/gene="[illegible]"
/note="[illegible]
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="[illegible]"
/protein_id="AA[illegible]"
/db_xref="[illegible]"
/translation="[illegible]"
gene complement([illegible])
/gene="[illegible]"
CDS complement([illegible])
/gene="[illegible]"
/note="[illegible]
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA[illegible]"
/db_xref="[illegible]"
/translation="[illegible]"
gene complement([illegible])
/gene="[illegible]"
CDS complement([illegible])
/gene="[illegible]"
/note="hypothetical protein; identified by [illegible];
putative"
/codon_start=1
/transl_table=11
/product="[illegible]"
/protein_id="AA[illegible]"
/db_xref="[illegible]"
/translation="MVKK[illegible]
[illegible]
EKVKK[illegible]NKN[illegible]IK"
gene complement([illegible])

```
                /gene="BB0675"
     CDS        complement(773.. 8602)
                /gene="BB0675"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BB0675"
                /protein_id="AAC67026.1"
                /db_xref="GI:2688[illegible]"
                /translation="[illegible]
                [illegible]
                ELSEIAKVNKW"
     gene       complement(8599..9261)
                /gene="BB0676"
     CDS        complement(8599..9261)
                /gene="BB0676"
                /note="similar to GB:L42023 SP:P44755 PID:1004013
                PID:1222592 PID:1294815 percent identity: 29.58;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="phosphoglycolate phosphatase (gph)"
                /protein_id="AAC67019.1"
                /db_xref="GI:2688602"
                /translation="MKIKA[illegible]
                [illegible]
                IK"
     gene       9504..11114

  Query Match             100.0%;  Score 21;  DB 1;  Length 13271;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGTTAAAAGT 21
                |||||||||||||||||||||
Db         75 ATGAAAAGGAAAGTTAAAAGT 95


RESULT 4
AC119739/c
LOCUS       AC119739               29158 bp    DNA     linear   PRI 21-AUG-2002
DEFINITION  Homo sapiens 3 BAC RP11-676G20 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC119739
VERSION     AC119739.5  GI:22380674
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 29158)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali Osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Arey,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,I., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brelen,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,L.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.I., Cothorn,E.F., Cavil,E.,
            Davila,M.L., Davis,C., Davy Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.L., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferragato,L., Flagg,N., Ford,J., Foster,P., Frantz,I.,
            Gabisi,A., Gao,Y., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,[illegible], Hodgson,A., Holder,M., Hollins,[illegible]
            Hollins,B., Homsi,F., Howard,S., Huber,[illegible], Hulyk,S., Hume,[illegible]
            Jackson,L.R., Jacobson,B., Jia,Y., Johnson,B., Jolivet,S.
            Jordah,S., Karlgren,E., Kelly,S., Khan,H., Kiel,[illegible], Kovar,[illegible]
            Kovar,C., Kratovil,[illegible], Kureshi,A., Landry,N., Leal,B., Lewis,[illegible]
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,[illegible]
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,[illegible], McLeod,M.P., Meador,M.,
            Mei,[illegible], Metzker,M., Miner,[illegible], Miner,[illegible], Mitchell,[illegible], Mohabbat,K.
            [illegible], Morgan,M., [illegible], [illegible], [illegible], Neal,[illegible]
            Nelson,[illegible], Newton,[illegible], Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            [illegible]
            Okwuonu,[illegible], [illegible]
            [illegible]riede,P., Parr,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            [illegible]
            Rojas,A., Rojubokan,I., Rolte,M., Ruiz,S., Savery,[illegible], Scherer,[illegible]
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., [illegible]
            Sonaike,I., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,[illegible], Taylor,[illegible]
            Taylor,I., Teltrod,R., Thomas,N., Thomas,S., [illegible]mani,K., Vasquez,[illegible]
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward Moore,S.,
            Warren,R., Washington,C., Wattington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,[illegible], Worley,K.[illegible], Wu,[illegible], Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.[illegible], Weinstock,G.[illegible] and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 29158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible] MAY 2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 29158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 29158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 21, 2002 this sequence version replaced gi:22094209.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. [illegible]) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Regions of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect [illegible]) to the
            EST and cDNA sequences. Genes [illegible]
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
```

[illegible] indicated in the [illegible]

[illegible] BASE [illegible] quality [illegible] bases. [illegible] base [illegible] can be found [illegible] /genbank [illegible]

[illegible]

[illegible] Mismatches [illegible] Gaps [illegible]

[illegible] ENV [illegible]

[illegible] repeat [illegible]

[illegible] Streptophyta; [illegible] phyta; Tracheophyta; [illegible] eudicots; [illegible] Solanum.

[illegible]

```
                     /note "[illegible]"
                     /rpt_type INVERTED
     repeat_region   [illegible]
                     /note "[illegible]"
                     /rpt_type INVERTED
     TATA_signal     [illegible]
     intron          [illegible]
     repeat_region   [illegible]
                     /note "[illegible]"
                     /rpt_type DIRECT
     repeat_region   [illegible]
                     /note "[illegible]"
                     /rpt_type DIRECT
     repeat_region   [illegible]
                     /note "[illegible]"
                     /rpt_type DIRECT
     repeat_region   [illegible]
                     /note "[illegible]"
                     /rpt_type DIRECT
     misc_feature    [illegible]
                     /note "branch point"
     CDS             [illegible]
                     /[illegible]
                     /[illegible] "[illegible]"
                     /[illegible] "[illegible]"
                     /db_xref "[illegible]"
                     /db_xref "[illegible]"
                     /translation "M[illegible]NVKAK[illegible]
                     [illegible]
                     [illegible]LEVESSD[illegible]NVKAK[illegible]
                     [illegible]
                     [illegible]LADYNIQKESILHEVIRLRGGD[illegible]"
     polyA_signal    3524..[illegible]
BASE COUNT     [illegible]
ORIGIN

  Query Match             [illegible]
  Best Local Similarity   [illegible]
  Matches   20;  Conservative  [illegible]  Mismatches  [illegible]  Indels  [illegible]  Gaps  [illegible]

Qy          1 AT[illegible]AAAA[illegible]AAA[illegible]AAAA[illegible]
Db         [illegible] AT[illegible]AAAA[illegible]AAA[illegible]AAAA[illegible]


RESULT [illegible]
AC[illegible]448/c
LOCUS       AC[illegible]448   [illegible] bp    DNA   [illegible]   HTG   [illegible]
DEFINITION  Rattus norvegicus clone [illegible] *** SEQUENCING IN PROGRESS
            ***, [illegible] ordered pieces.
ACCESSION   AC[illegible]448
VERSION     AC[illegible]448.[illegible]
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   [illegible] Adams,[illegible]
            [illegible]
            Barbara,[illegible] Benton,[illegible]
            [illegible]
            [illegible] Burch,[illegible] Burkett,[illegible] Byrd,N.[illegible]
            Carron,[illegible] Carter,M.,[illegible]
            Chen,[illegible] Chen,R., Chen,[illegible] Chowdhury,[illegible] Christopoulos,[illegible]
            Cleveland,[illegible] Cox,[illegible]
            [illegible] Davis,[illegible] [illegible]
```

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,S., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogoh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,I., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 96637)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17 SEP 2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 96637)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11 FEB 2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 10, 2002 this sequence version replaced gi:17950650.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.[illegible]/
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GFTA
Center clone name: CH230-129E12

Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye Terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 40976 bases at least Q40
Consensus quality: 44328 bases at least Q30
Consensus quality: 46474 bases at least Q20
----

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.[illegible]s/Clone_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1618: contig of 1618 bp in length
* 1619 1718: gap of unknown length
* 1719 3509: contig of 1791 bp in length
* 3510 3609: gap of unknown length
* 3610 4670: contig of 1061 bp in length
* 4671 4770: gap of unknown length
* 4771 5908: contig of 1138 bp in length
* 5909 6008: gap of unknown length
* 6009 7170: contig of 1162 bp in length
* 7171 7270: gap of unknown length
* 7271 9009: contig of 1739 bp in length
* 9010 9109: gap of unknown length
* 9110 10472: contig of 1363 bp in length
* 10473 10572: gap of unknown length
* 10573 11603: contig of 1031 bp in length
* 11604 11703: gap of unknown length
* 11704 13004: contig of 1301 bp in length
* 13005 13104: gap of unknown length
* 13105 14269: contig of 1165 bp in length
* 14270 14369: gap of unknown length
* 14370 15814: contig of 1445 bp in length
* 15815 15914: gap of unknown length
* 15915 16970: contig of 1056 bp in length
* 16971 17070: gap of unknown length
* 17071 18661: contig of 1591 bp in length
* 18662 18761: gap of unknown length
* 18762 20624: contig of 1863 bp in length
* 20625 20724: gap of unknown length
* 20725 21940: contig of 1216 bp in length
* 21941 22040: gap of unknown length
* 22041 23446: contig of 1406 bp in length
* 23447 23546: gap of unknown length
* 23547 25328: contig of 1782 bp in length
* 25329 25428: gap of unknown length
* 25429 26526: contig of 1098 bp in length
* 26527 26626: gap of unknown length
* 26627 27892: contig of 1266 bp in length
* 27893 27992: gap of unknown length
* 27993 29004: contig of 1012 bp in length
* 29005 29104: gap of unknown length
* 29105 30756: contig of 1652 bp in length
* 30757 30856: gap of unknown length
* 30857 32078: contig of 1222 bp in length
* 32079 32178: gap of unknown length
* 32179 34028: contig of 1850 bp in length
* 34029 34128: gap of unknown length
* 34129 35241: contig of 1113 bp in length
* 35242 35341: gap of unknown length
* 35342 36460: contig of 1119 bp in length
* 36461 36560: gap of unknown length
* 36561 38024: contig of 1464 bp in length
* 38025 38124: gap of unknown length
* 38125 39517: contig of 1393 bp in length
* 39518 39617: gap of unknown length
* 39618 41946: contig of 2329 bp in length
* 41947 42046: gap of unknown length
* 42047 43709: contig of 1663 bp in length
* 43710 43809: gap of unknown length
* 43810 45157: contig of 1348 bp in length
* 45158 45257: gap of unknown length
* 45258 47214: contig of 1957 bp in length
* 47215 47314: gap of unknown length
* 47315 49842: contig of 2528 bp in length
* 49843 49942: gap of unknown length
* 49943 51895: contig of 1953 bp in length
* 51896 51995: gap of unknown length
* 51996 53689: contig of 1694 bp in length
* 53690 53789: gap of unknown length
* 53790 55656: contig of 1867 bp in length
* 55657 55756: gap of unknown length
* 55757 57239: contig of 1483 bp in length
* 57240 57339: gap of unknown length
* 57340 58891: contig of 1552 bp in length
* 58892 58991: gap of unknown length

[illegible]

*** [illegible] HTG [illegible] 2002
[illegible] *** [illegible]NTING IN PROGRESS

[illegible] Mammalian Gene Collection [illegible]
[illegible] Machine; Mammal[illegible]

[illegible] Allen,[illegible]
[illegible] Banks,[illegible]
[illegible] Bonnin,[illegible]
[illegible] Bryant,N.[illegible]
[illegible] Byrd,N.[illegible]
[illegible] Chavez,[illegible]
[illegible] Edwards,[illegible] Escotto,M.,
[illegible] Hamilton,K.,
[illegible] Hart,M.[illegible] Hernandez,[illegible]

Hernandez,[illegible] Hodgson,A., [illegible]
Hulse[illegible] Howard,S., [illegible]
Jacobsen,[illegible]
Karlsson,F., [illegible] Khan,[illegible]
[illegible]
Li,[illegible] Li,Z., Lichtarge,[illegible] Townsend,H.[illegible]
[illegible] Ma[illegible]
Maheshwari,M., Majesa,[illegible] Martin,[illegible] Martinez,A., Martinez,[illegible]
Mangey,[illegible] Mawhiney,[illegible]
Miner,[illegible] Miner,[illegible] Mitchell,[illegible]
Moser,M., Neal,[illegible] Newtson,[illegible] Newton,N., Nguyen,A., Nguyen,N.,
Ngoyen,N., Nickerson,[illegible]
Okwuonu,[illegible]
Odeaunye,N., Oviedo,R., Pace,A., Payton,[illegible] Perry,[illegible]
Peters,[illegible]
Rives,M., Rojas,A., [illegible] Savery,[illegible]
Scherer,S., Scott,[illegible] Shen,H., [illegible]
Sodergren,[illegible] Sparks,A., [illegible]
Sutton,A., Svatek,A., Tabor,[illegible]
Tansey,[illegible]
Usmani,K., Vasquez,[illegible] Vera,V., Villasana,[illegible] Wang,[illegible]
Wang,S., Ward[illegible] Warren,[illegible]
Williams,[illegible] Williamson,A., [illegible] Worley,K.
Wu,[illegible] Wu,Y., Wu,Y.[illegible]
Weinstock,[illegible] and Gibbs,R.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to [illegible])
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (22 FEB 2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX [illegible], USA
REFERENCE	3 (bases 1 to [illegible])
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted ([illegible] 2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX [illegible], USA
COMMENT	On Jul 14, 2002 this sequence version replaced [illegible]

[illegible] Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu
Contact: [illegible]
[illegible]
Center project name: [illegible]
Center clone name: [illegible]
[illegible] Summary Statistics
Sequencing vector: [illegible]
Chemistry: [illegible]
Assembly program: Phrap; version [illegible]
Consensus quality: [illegible] bases at least [illegible]
Consensus quality: [illegible] bases at least [illegible]
Consensus quality: [illegible] bases at least [illegible]

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/[illegible]).
* NOTE: This is a 'working draft' sequence. It currently
* consists of [illegible] contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* [illegible] contig of [illegible] bp in length
* 1476 [illegible] gap of unknown length
* [illegible] contig of [illegible] bp in length
* [illegible] gap of unknown length
* [illegible] contig of [illegible] bp in length
* [illegible] gap of unknown length
* 4412 [illegible] contig of [illegible] bp in length
* [illegible] gap of unknown length

```
*     5580      6968: contig of 1389 bp in length
*     6969      7068: gap of unknown length
*     7069      8810: contig of 1742 bp in length
*     8811      8910: gap of unknown length
*     8911     10608: contig of 1698 bp in length
*    10609     10708: gap of unknown length
*    10709     11826: contig of 1118 bp in length
*    11827     11926: gap of unknown length
*    11927     12928: contig of 1002 bp in length
*    12929     13028: gap of unknown length
*    13029     14989: contig of 1961 bp in length
*    14990     15089: gap of unknown length
*    15090     16873: contig of 1784 bp in length
*    16874     16973: gap of unknown length
*    16974     18382: contig of 1409 bp in length
*    18383     18482: gap of unknown length
*    18483     20350: contig of 1868 bp in length
*    20351     20450: gap of unknown length
*    20451     21831: contig of 1381 bp in length
*    21832     21931: gap of unknown length
*    21932     23379: contig of 1448 bp in length
*    23380     23479: gap of unknown length
*    23480     24602: contig of 1123 bp in length
*    24603     24702: gap of unknown length
*    24703     26187: contig of 1485 bp in length
*    26188     26287: gap of unknown length
*    26288     27577: contig of 1290 bp in length
*    27578     27677: gap of unknown length
*    27678     28897: contig of 1220 bp in length
*    28898     28997: gap of unknown length
*    28998     31341: contig of 2344 bp in length
*    31342     31441: gap of unknown length
*    31442     33744: contig of 2303 bp in length
*    33745     33844: gap of unknown length
*    33845     35113: contig of 1269 bp in length
*    35114     35213: gap of unknown length
*    35214     37433: contig of 2220 bp in length
*    37434     37533: gap of unknown length
*    37534     40687: contig of 3154 bp in length
*    40688     40787: gap of unknown length
*    40788     42386: contig of 1599 bp in length
*    42387     42486: gap of unknown length
*    42487     44948: contig of 2462 bp in length
*    44949     45048: gap of unknown length
*    45049     47757: contig of 2709 bp in length
*    47758     47857: gap of unknown length
*    47858     52063: contig of 4206 bp in length
*    52064     52163: gap of unknown length
*    52164     55933: contig of 3770 bp in length
*    55934     56033: gap of unknown length
*    56034     59637: contig of 3604 bp in length
*    59638     59737: gap of unknown length
*    59738     63980: contig of 4243 bp in length
*    63981     64080: gap of unknown length
*    64081     68377: contig of 4297 bp in length
*    68378     68477: gap of unknown length
*    68478     71796: contig of 3319 bp in length
*    71797     71896: gap of unknown length
*    71897     75280: contig of 3384 bp in length
*    75281     75380: gap of unknown length
*    75381     79957: contig of 4577 bp in length
*    79958     80057: gap of unknown length
*    80058     84069: contig of 4012 bp in length
*    84070     84169: gap of unknown length
*    84170     91648: contig of 7479 bp in length
*    91649     91748: gap of unknown length
*    91749     98922: contig of 7174 bp in length
*    98923     99022: gap of unknown length
*    99023    108998: contig of 9976 bp in length
*   108999    109098: gap of unknown length
*   109099    119981: contig of 10883 bp in length
*   119982    120081: gap of unknown length
*   120082    129933: contig of 9852 bp in length
*   129934    130033: gap of unknown length
*   130034    143565: contig of 13532 bp in length
*   143566    143665: gap of unknown length
*   143666    157175: contig of 13510 bp in length
FEATURES             Location/Qualifiers
     source          1..157175
                     /organism "Rattus norvegicus"
                     /db_xref "taxon:10116"
                     /clone "CH230-296C16"
BASE COUNT     49619 a  32579 c  33537 g  39996 t   1,004 others
ORIGIN

  Query Match             95.2%;  Score 20;  DB 1;  Length 157175;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGAAAAGAAAGTAAAAG 20
                |||||||||||||||||
Db     154797 AGAAAAGAAAGTAAAAG 154816


RESULT 8
AC100907/c
LOCUS       AC100907               54245 bp    DNA     linear   HTG 25-NOV-2001
DEFINITION  Mus musculus clone RP23-7012, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100907
VERSION     AC100907.1  GI:17059681
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 54245)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-7012
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 54245)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abi,A., Allen,N.
            Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.
            Choepel,Y., Colangelo,M., Collins,S., Corlymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Hinde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaBocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogers,J.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., St Jane,V.,N.
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
                ----- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
                        - Project Information
                Center project name: 114548
                Center clone name: 7012_2
```

[illegible]

FEATURES [illegible]
source [illegible]
/organism="[illegible]"
/[illegible]="[illegible]"
/[illegible]="[illegible]"
/[illegible]="[illegible]"
BASE COUNT [illegible]

```
ORIGIN

  Query Match             92.4%;  Score 19.4;  DB 2;  Length 54245;
  Best Local Similarity   95.2%;  Pred. No. 3.1e+02;
  Matches   20;  Conservative [illegible];  Mismatches  1;  Indels [illegible];  Gaps [illegible];

Qy          1 ATGAAAAGGAAAGCTAAAAGT 21
              |||||||||||||||||||||
Db      42590 ATGGAAAGGAAAGCTAAAAGT 42570


RESULT 9
AC092636/c
LOCUS       AC092636               199721 bp    DNA     linear   PRI 16-APR-2002
DEFINITION  Homo sapiens BAC clone RP11-353K11 from 2, complete sequence.
ACCESSION   AC092636 AC068117
VERSION     AC092636.3  GI:19807915
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199721)
  AUTHORS   Dalston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 199721)
  AUTHORS   Ali,J., Cotton,M. and Spalding,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-353K11
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 199721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 199721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 199721)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 29, 2002 this sequence version replaced gi:15624963.
                        Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://[illegible].wustl.edu/[illegible]
               Contact: sapiens@watson.wustl.edu
                        Summary Statistics
               Center project name: H_NH0353K11
               Drafting Center: WIBR

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by [illegible]
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 52:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR: pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-14[illegible]M[illegible]; the clone sequenced
            to the right is RP11-219H23. Actual start of this clone is at base
            position 1 of RP11-353K11; actual end is at base position [illegible] of
            RP11-353K11.

            The region from 147655 to 147703; [illegible] are only covered
            by pcr products from clone DNA. Single subclone region exist
            between 147951 and 147966. Polymorphisms exist between A[illegible] and
            AC092636. Data from AC068[illegible] and AC07[illegible] was used to finish
            AC092636.

            The sequence of AC068117 has been incorporated into AC092636.
FEATURES             Location/Qualifiers
     source          1..199721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-353K11"
                     /clone_lib="RPCI-11"
     repeat_region   1..1237
                     /rpt_family="Alu"
     repeat_region   1066..1248
                     /rpt_family="MIR"
     repeat_region   1256..1558
                     /rpt_family="Alu"
     repeat_region   1559..1829
                     /rpt_family="L2"
     repeat_region   1830..2142
                     /rpt_family="Alu"
     repeat_region   2143..2327
                     /rpt_family="L2"
     repeat_region   2473..2757
                     /rpt_family="Alu"
     repeat_region   2761..3070
                     /rpt_family="Alu"
     repeat_region   3145..3649
                     /rpt_family="MaLR"
     repeat_region   3653..3954
                     /rpt_family="Alu"
     repeat_region   4149..4414
                     /rpt_family="Alu"
     repeat_region   4833..4858
                     /rpt_family="(TTTG)n"
     repeat_region   4867..5177
                     /rpt_family="Alu"
     repeat_region   5178..5212
                     /rpt_family="(TTTTG)n"
     repeat_region   5213..5469
                     /rpt_family="Alu"
     repeat_region   5650..5818
                     /rpt_family="Alu"
     misc_feature    5650..5665
                     /note="match to ESI AI[illegible] (NIH:a44[illegible]) tjt[illegible]x1"
     misc_feature    5650..5665
                     /note="match to ESI AI[illegible] (NIH:a44[illegible]) t[illegible]x1"
```

[illegible]

"A[illegible]"

"[illegible]"

"M[illegible]"

"A[illegible]"

[illegible]

50087 ATGAAAAGAAAAGAAAA[illegible]

RESULT 10
A[illegible]
LOCUS A[illegible] DNA [illegible]
DEFINITION Rattus [illegible] *** [illegible] IN [illegible] *** [illegible]
ACCESSION A[illegible]
VERSION A[illegible]
KEYWORDS HTG; [illegible] PHASE [illegible]
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 157847)
AUTHORS Murphy,[illegible]M., Adams,[illegible], A[illegible], Alsbrooks,[illegible], Amaratunga,[illegible], Barbara,[illegible], Bowles,[illegible], Brown,M., Brown,[illegible], Burkett,[illegible], Carter,M., [illegible], Davila,M.[illegible], Davis,[illegible], Delaney,K.[illegible], Draper,H.[illegible], Edwards,[illegible], Falls,[illegible], Ferraguto,[illegible], Gabisi,A., Gao,[illegible], Garcia,A., Garner,[illegible], Gorrell,J.H., Guevara,W., Gunaratne,[illegible], Hamilton,K., Harris,[illegible], Harris,K., Hart,M., Havlak,[illegible], Hawes,A., Hernandez,[illegible], Hernandez,[illegible], Hodgson,A., [illegible], Hulyk,[illegible], Jackson,[illegible], Jacobs,[illegible], Karlsson,[illegible], Khan,[illegible], Kravitz,[illegible], Kowis,[illegible], Lewis,[illegible], Maheshwari,M., Major,[illegible], Martin,[illegible], Martinez,[illegible], Mawhiney,[illegible], Miner,[illegible], Morgan,M., Moser,M., Neal,[illegible], Newton,[illegible], Nguyen,[illegible], Nguyen,N., Nickerson,E., Nwokenkwo,[illegible], Okwuonu,[illegible], Perkins,[illegible], Peters,[illegible], Pickens,[illegible], Rives,M., Rojas,A., [illegible], Scherer,S., Scott,[illegible], Shen,H., [illegible], Sodergren,[illegible], Sparks,A., [illegible], Sutton,A., Svatek,A., [illegible], Usmani,K., Vasquez,[illegible], Vera,V., [illegible], Ward,[illegible], Ward-Moore,S., [illegible], Williams,[illegible], Williamson,A., [illegible], Worley,K.[illegible] and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to [illegible])
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (25 NOV 2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, [illegible] Baylor Plaza, Houston, TX [illegible], USA
REFERENCE 3 (bases 1 to 157847)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted ([illegible]) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, [illegible] Baylor Plaza, Houston, TX [illegible], USA
COMMENT [illegible]

```
             Genome Center
       Center: Baylor College of Medicine
       Center code: BCM
       Web site: http://www.hgsc.bcm.tmc.edu/
       Contact: hgsc-help@bcm.tmc.edu
       ----       Project Information
       Center project name: GHUH
       Center clone name: CH230-102H5
       --         Summary Statistics
       Sequencing vector: Plasmid;
       Chemistry: Dye-terminator Big Dye; 100% of reads
       Assembly program: Phrap; version 0.990329
       Consensus quality: 108897 bases at least Q40
       Consensus quality: 114669 bases at least Q30
       Consensus quality: 119448 bases at least Q20
       ----
     * NOTE: Estimated insert size may differ from sequence length
     *   (see http://www.[illegible]s/Genbank_draft_data.html).
     * NOTE: This is a 'working draft' sequence. It currently
     * consists of 67 contigs. The true order of the pieces
     * is not known and their order in this sequence record is
     * arbitrary. Gaps between the contigs are represented as
     * runs of N, but the exact sizes of the gaps are unknown.
     * This record will be updated with the finished sequence
     * as soon as it is available and the accession number will
     * be preserved.
     *          1      1244: contig of 1244 bp in length
     *       1245      1344: gap of unknown length
     *       1345      2472: contig of 1128 bp in length
     *       2473      2572: gap of unknown length
     *       2573      3983: contig of 1411 bp in length
     *       3984      4083: gap of unknown length
     *       4084      5626: contig of 1543 bp in length
     *       5627      5726: gap of unknown length
     *       5727      7141: contig of 1415 bp in length
     *       7142      7241: gap of unknown length
     *       7242      8484: contig of 1243 bp in length
     *       8485      8584: gap of unknown length
     *       8585      9946: contig of 1362 bp in length
     *       9947     10046: gap of unknown length
     *      10047     11078: contig of 1032 bp in length
     *      11079     11178: gap of unknown length
     *      11179     12465: contig of 1287 bp in length
     *      12466     12565: gap of unknown length
     *      12566     14077: contig of 1512 bp in length
     *      14078     14177: gap of unknown length
     *      14178     15568: contig of 1391 bp in length
     *      15569     15668: gap of unknown length
     *      15669     16852: contig of 1184 bp in length
     *      16853     16952: gap of unknown length
     *      16953     18216: contig of 1264 bp in length
     *      18217     18316: gap of unknown length
     *      18317     19432: contig of 1116 bp in length
     *      19433     19532: gap of unknown length
     *      19533     21624: contig of 2092 bp in length
     *      21625     21724: gap of unknown length
     *      21725     23299: contig of 1575 bp in length
     *      23300     23399: gap of unknown length
     *      23400     24897: contig of 1498 bp in length
     *      24898     24997: gap of unknown length
     *      24998     26011: contig of 1014 bp in length
     *      26012     26111: gap of unknown length
     *      26112     27759: contig of 1648 bp in length
     *      27760     27859: gap of unknown length
     *      27860     29047: contig of 1188 bp in length
     *      29048     29147: gap of unknown length
     *      29148     30315: contig of 1168 bp in length
     *      30316     30415: gap of unknown length
     *      30416     31600: contig of 1185 bp in length
     *      31601     31700: gap of unknown length
     *      31701     33358: contig of 1658 bp in length
     *      33359     33458: gap of unknown length
     *      33459     35317: contig of 1859 bp in length
     *      35318     35417: gap of unknown length
     *      35418     37013: contig of 1596 bp in length
     *      37014     37113: gap of unknown length
     *      37114     39445: contig of 2332 bp in length
     *      39446     39545: gap of unknown length
     *      39546     40784: contig of 1239 bp in length
     *      40785     40884: gap of unknown length
     *      40885     42363: contig of 1479 bp in length
     *      42364     42463: gap of unknown length
     *      42464     44470: contig of 2007 bp in length
     *      44471     44570: gap of unknown length
     *      44571     46177: contig of 1607 bp in length
     *      46178     46277: gap of unknown length
     *      46278     47377: contig of 1100 bp in length
     *      47378     47477: gap of unknown length
     *      47478     49577: contig of 2100 bp in length
     *      49578     49677: gap of unknown length
     *      49678     50739: contig of 1062 bp in length
     *      50740     50839: gap of unknown length
     *      50840     52819: contig of 1980 bp in length
     *      52820     52919: gap of unknown length
     *      52920     53931: contig of 1012 bp in length
     *      53932     54031: gap of unknown length
     *      54032     55170: contig of 1139 bp in length
     *      55171     55270: gap of unknown length
     *      55271     57361: contig of 2091 bp in length
     *      57362     57461: gap of unknown length
     *      57462     59521: contig of 2060 bp in length
     *      59522     59621: gap of unknown length
     *      59622     60984: contig of 1363 bp in length
     *      60985     61084: gap of unknown length
     *      61085     63052: contig of 1968 bp in length
     *      63053     63152: gap of unknown length
     *      63153     66088: contig of 2936 bp in length
     *      66089     66188: gap of unknown length
     *      66189     69422: contig of 3234 bp in length
     *      69423     69522: gap of unknown length
     *      69523     72422: contig of 2900 bp in length
     *      72423     72522: gap of unknown length
     *      72523     76051: contig of 3529 bp in length
     *      76052     76151: gap of unknown length
     *      76152     79079: contig of 2928 bp in length
     *      79080     79179: gap of unknown length
     *      79180     81264: contig of 2085 bp in length
     *      81265     81364: gap of unknown length
     *      81365     83968: contig of 2604 bp in length
     *      83969     84068: gap of unknown length
     *      84069     87421: contig of 3353 bp in length
     *      87422     87521: gap of unknown length
     *      87522     89433: contig of 1912 bp in length
     *      89434     89533: gap of unknown length
     *      89534     92806: contig of 3273 bp in length
     *      92807     92906: gap of unknown length
     *      92907     97145: contig of 4239 bp in length
     *      97146     97245: gap of unknown length
     *      97246    100195: contig of 2950 bp in length
     *     100196    100295: gap of unknown length
     *     100296    102907: contig of 2612 bp in length
     *     102908    103007: gap of unknown length
     *     103008    104887: contig of 1880 bp in length

  Query Match             90.5%;  Score 19;  DB [illegible];  Length [illegible];
  Best Local Similarity   100.0%;  Pred. No. [illegible];
  Matches   19;  Conservative    0;  Mismatches    [illegible];  Indels    [illegible];  Gaps    [illegible];

Qy          1 ATGAAAAGGAAAGCTAAAA 19
                |||||||||||||||||||
Db      58423 ATGAAAAGGAAAGCTAAAA 58441


RESULT 11
AC126997
LOCUS       AC126997                174417 bp    DNA     linear   HTG 14-[illegible]-2002
```

[illegible] SEQUENCING IN PROGRESS

[illegible]

[illegible]

* NOTE: [illegible]
* (see http://www.[illegible])
* NOTE: This is a 'working draft' sequence. It currently
* consists of [illegible] contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* [illegible]

*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	10496: contig of 1464 bp in length
*	10497	10696: gap of unknown length
*	10697	[illegible]: contig of [illegible] bp in length
*	12202	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length
*	[illegible]	[illegible]: contig of [illegible] bp in length
*	[illegible]	[illegible]: gap of unknown length

```
  *     44018     45162: contig of 1145 bp in length
  *     45163     45262: gap of unknown length
  *     45263     47031: contig of 1769 bp in length
  *     47032     47131: gap of unknown length
  *     47132     48327: contig of 1196 bp in length
  *     48328     48427: gap of unknown length
  *     48428     49509: contig of 1082 bp in length
  *     49510     49609: gap of unknown length
  *     49610     51952: contig of 2343 bp in length
  *     51953     52052: gap of unknown length
  *     52053     53598: contig of 1546 bp in length
  *     53599     53698: gap of unknown length
  *     53699     55695: contig of 1997 bp in length
  *     55696     55795: gap of unknown length
  *     55796     57128: contig of 1333 bp in length
  *     57129     57228: gap of unknown length
  *     57229     58733: contig of 1505 bp in length
  *     58734     58833: gap of unknown length
  *     58834     60074: contig of 1241 bp in length
  *     60075     60174: gap of unknown length
  *     60175     61805: contig of 1631 bp in length
  *     61806     61905: gap of unknown length
  *     61906     63424: contig of 1519 bp in length
  *     63425     63524: gap of unknown length
  *     63525     64734: contig of 1210 bp in length
  *     64735     64834: gap of unknown length
  *     64835     66248: contig of 1414 bp in length
  *     66249     66348: gap of unknown length
  *     66349     69187: contig of 2839 bp in length
  *     69188     69287: gap of unknown length
  *     69288     71026: contig of 1739 bp in length
  *     71027     71126: gap of unknown length
  *     71127     73420: contig of 2294 bp in length
  *     73421     73520: gap of unknown length
  *     73521     74785: contig of 1265 bp in length
  *     74786     74885: gap of unknown length
  *     74886     77510: contig of 2625 bp in length
  *     77511     77610: gap of unknown length
  *     77611     79407: contig of 1797 bp in length
  *     79408     79507: gap of unknown length
  *     79508     82288: contig of 2781 bp in length
  *     82289     82388: gap of unknown length
  *     82389     83736: contig of 1348 bp in length
  *     83737     83836: gap of unknown length
  *     83837     85986: contig of 2150 bp in length
  *     85987     86086: gap of unknown length
  *     86087     88381: contig of 2295 bp in length
  *     88382     88481: gap of unknown length
  *     88482     90888: contig of 2407 bp in length
  *     90889     90988: gap of unknown length
  *     90989     92994: contig of 2006 bp in length
  *     92995     93094: gap of unknown length
  *     93095     95359: contig of 2265 bp in length

  Query Match             90.5%;  Score 19;  DB 2;  Length 171417,
  Best Local Similarity   100.0%;  Pred. No. 3.9e+02,
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 ATGAAAAGGAAAGCTAAAA 19
            |||||||||||||||||||
Db   124653 ATGAAAAGGAAAGCTAAAA 124671


RESULT 12
CEL25C12/
LOCUS       CEL25C12            33045 bp    DNA     linear   INV 24-JAN-2002
DEFINITION  Caenorhabditis elegans cosmid T25C12, complete sequence.
ACCESSION   Z66566
VERSION     Z66566.1  GI:1051339
KEYWORDS    HTG; Lin 14.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
   PUBMED   9851916
  REMARK    The C.elegans Sequencing Consortium
REFERENCE   2  (bases 1 to 33045)
  AUTHORS   Kershaw,J.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04 NOV 1995) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            This sequence is the entire insert of clone T25C12. The true right
            end of clone F48C5 is at 1899 in this sequence. The start of this
            sequence (1..105) overlaps with the end of sequence Z68107.
            The end of this sequence (32947..33045) overlaps with the start of
            sequence Z93779.
            For a graphical representation of this sequence and its analysis
            see:- http://www.sanger.ac.uk/cgi-bin/ace/simple/Sequence?
            name=T25C12
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
FEATURES             Location/Qualifiers
     source          1..33045
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="X"
                     /clone="T25C12"
     gene            join(9132..9253,20428..20645,21083..21286,22191..22270,
                     22753..22929,22980..23088,23135..23350,23398..23563,
                     23615..23762,23807..23914)
                     /gene="T25C12.1a"
     CDS             join(9132..9253,20428..20645,21083..21286,22191..22270,
                     22753..22929,22980..23088,23135..23350,23398..23563,
                     23615..23762,23807..23914)
                     /gene="T25C12.1a"
                     /note="Embryonic nuclear protein lin-14 form B2
                     cDNA EST yk260a11.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA91484.1"
                     /db_xref="GI:2990223"
                     /db_xref="SWISS-PROT:Q21446"
                     /translation="MNI [illegible] KEVI
                     DLRQAVNLILPMLPLYPTIGNGPNAT[illegible]LAAQPT[illegible]HVIQQSLLKKRPVAQ[illegible]VPQPE
                     CPGQIKPVLSSPAAALQNVIMLNPWIMG[illegible]LKPAS[illegible]LFNGQIP[illegible]LIGETSLQGIDDQ
                     TVKWIGPSSVDSNGQKTDSSAASAGDNQNIDVIG[illegible]SPTG[illegible]NH[illegible]AQ[illegible]IAL[illegible]ISQQT
                     F[illegible]NALKDSS[illegible]INPVPIVETAP[illegible]VAP[illegible]IN[illegible]IN[illegible]AKA[illegible]FEKKPK[illegible]VN[illegible]IVE
                     IVRNQDLSEENISMFEIPVPKAIASD[illegible]IFRPVSE[illegible]IQQIIQGKKYEEMEVGE[illegible]MIQ
                     LCKRLA[illegible]KEV[illegible]MS[illegible]TTVA[illegible]NE[illegible]NVAN[illegible]HV[illegible]EV[illegible]NEEDF
                     WDKFREAMRKLAAR[illegible]KEVKHAKK[illegible]NKEEAQALML[illegible]KKYG[illegible]M[illegible]N[illegible]NGA[illegible]EV
                     [illegible]VSEFAN[illegible]F[illegible]QNI[illegible]SQS
                     PPLQGPQ"
     gene            join(20139..20326,20428..20645,21083..21286,22191..22270,
                     22753..22929,22980..23088,23135..23350,23398..23563,
                     23615..23762,23807..23914)
                     /gene="T25C12.1b"
     CDS             join(20139..20326,20428..20645,21083..21286,22191..22270,
                     22753..22929,22980..23088,23135..23350,23398..23563,
                     23615..23762,23807..23914)
```

[illegible] form A"
[illegible]
/translation="[illegible]MRQAAL[illegible]
[illegible]MSAFRKV[illegible]NA[illegible]AAL[illegible]
[illegible]VALIPIVI[illegible]NVIMLNPWIMGSSLKPASP
[illegible]SSAASA[illegible]
[illegible]AIMI[illegible]IVLIA[illegible]LPVAP[illegible]
[illegible]KPVN[illegible]VKIV[illegible]PVPKAIASDPIPRPVSEQQ
[illegible]IVA[illegible]NHSNYAN[illegible]
[illegible]KPENP[illegible]VKHAKKIKHNKKEAQAPML
[illegible]IKSQLESLPAHIPKIESEIPLI
[illegible]KIKVFS[illegible]"
[illegible](141)
[illegible](141)
[illegible]
/translation="MKIV[illegible]PLVAVVAV[illegible]AIAP[illegible]AVN[illegible]
[illegible]KKE[illegible]IKEIES[illegible]MKIVL
[illegible]PAAA[illegible]AVK[illegible]PPHIAP[illegible]AP[illegible]KKEIK
[illegible]SHSPP[illegible]IMP[illegible]V[illegible]"
[illegible]

complement([illegible])
complement([illegible]),complement([illegible]))
complement([illegible]),complement([illegible])
complement([illegible]),complement([illegible])
complement([illegible]),complement([illegible])
/gene="[illegible]"
/note="contains similarity to [illegible]
[illegible]"
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST yk[illegible]
comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST
yk[illegible] comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST
yk[illegible] comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST
yk[illegible] comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST
yk[illegible] comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST
yk[illegible] comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST
yk[illegible] comes from this gene
cDNA EST yk[illegible] comes from this gene; cDNA EST

Query Match [illegible]
Best Local Similarity [illegible]
Matches 19; Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

Qy 1 ATGAAAA[illegible]GAAA[illegible]AAAA[illegible]
Db [illegible]344 ATGAAAA[illegible]GAAAA[illegible]AAAA[illegible]

RESULT 13
A[illegible]344/c
LOCUS A[illegible]344 [illegible] DNA [illegible] HTG [illegible]
DEFINITION Homo sapiens chromosome [illegible] WORKING DRAFT
SEQUENCE, 4 unordered pieces.
ACCESSION A[illegible]344
VERSION A[illegible]344[illegible]
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to [illegible]344[illegible])
AUTHORS Murray,[illegible]M., Adams,[illegible], A[illegible], [illegible]
Alstrom[illegible]S.[illegible], Amatruda[illegible], [illegible]
Barbara[illegible], Bent[illegible], [illegible]
Boeck[illegible], Bowie,S., Briceno,M., Brown,[illegible], Brown,M., Bryant,N.[illegible]

```
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,I.F., Carter,M., Cazazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,K., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,F., Kelly,S., Khan,U., King,L., Korvih,L., Kovar,C.,
            Kratovic,L., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Liou,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.I., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,C., Miner,Z., Mitchell,L., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwaokelemeh,O., Nwokwuh,M., Nwokwu,C.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Perry,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojabokan,I., Ruiz,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,O., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward Moore,J., Warren,R., Washington,C., Watlington,S.,
            Williams,T., Williamson,A., Wleczyk,P., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133445)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18 JUN 2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 133445)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09 MAY 2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2002 this sequence version replaced gi:16117941.
               ------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
                      ------- Project Information
               Center project name: HBCF
               Center clone name: RP11-100E14
                      ------- Summary Statistics
               Sequencing vector: M13;
               Chemistry: Dye-primer Bodipy: 19% of reads
               Chemistry: Dye-terminator Big Dye: 81% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 132033 bases at least Q40
               Consensus quality: 135474 bases at least Q30
               Consensus quality: 137320 bases at least Q20
               Estimated insert size: 134201; sum-of-contigs estimation
               Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
               -------
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      2013: contig of 2013 bp in length
            *     2014      2113: gap of unknown length
            *     2114     41789: contig of 39676 bp in length
            *    41790     41889: gap of unknown length
            *    41890     79584: contig of 37695 bp in length
            *    79585     79684: gap of unknown length
            *    79685    133445: contig of 53761 bp in length
FEATURES             Location/Qualifiers
     source          1..133445
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-100E14"
BASE COUNT    33186 a  25703 c  25551 g  34905 t     ... others
ORIGIN

  Query Match             97.9%;  Score 19.6;  DB 1;  Length 133445;
  Best Local Similarity   95.0%;  Pred. No. 7e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps

Qy          2 TCAAAAGGAAAGCTAAAAGT 21
              | ||||| ||||||||||||
Db      94269 TAAAAAGAAAGCTAAAAGT 94250


RESULT 14
AC011421
LOCUS       AC011421              148142 bp    DNA     linear   PRI 19-APR-2000
DEFINITION  Homo sapiens chromosome 5 clone CIT-HSPC..., complete sequence.
ACCESSION   AC011421
VERSION     AC011421.2  GI:6102634
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148142)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148142)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 148142)
  AUTHORS   DOE Joint Genome Center and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-1999) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 148142)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 148142)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   6  (bases 1 to 148142)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 23, 1999 this sequence version replaced gi:6015121.
```

Search completed: November ...

[illegible] Seconds
[illegible] updates/se[illegible]

[illegible] summaries

	12	17.4			24	ABK[illegible]	cDNA encoding human
	17	17.4	82.9	34712		ABL[illegible]	Drosophila melanog
	21	16.8		983	22	ABA4[illegible]	Human breast [illegible]
	25	16.8	80.0	983	22	AAI[illegible]	Probe #4100 used [illegible]
	29	16.8	80.0	476		ABA[illegible]	Human [illegible]
	33	16.8	80.0	476		AA[illegible]44	Probe [illegible] used [illegible]
c	37	16.8	80.0	654	24	ABK[illegible]	Human cDNA [illegible]
	40	16.8	80.0			AAX4[illegible]	[illegible] cancer asso
c	41	16.8	80.0	2583	22	AAS[illegible]445	Human genomic DNA
	42	16.8	80.0	[illegible]46	22	AAS[illegible]	DNA encoding [illegible]
	43	16.8	80.0	[illegible]46		ABL[illegible]	Drosophila melanog
c	44	16.8	80.0	[illegible]421	24	ABL[illegible]	Lung cancer [illegible]
c	45	16.8	80.0	[illegible]421	24	ABL[illegible]	Kidney cancer [illegible]

ALIGNMENTS

RESULT 1
AAX89521
ID AAX89521 standard; DNA; 20 BP.
XX
AC AAX89521;
XX
DT [illegible] 1999 (first entry)
XX
DE PCR primer [illegible]
XX
KW [illegible]; FlaA; Lyme disease; antigen; assay; PCR primer; [illegible]
XX
OS Synthetic.
OS Borrelia burgdorferi.
XX
PN W0999[illegible]-A1.
XX
PD 15-JUL-1999.
XX
PF [illegible]-JAN-1999; [illegible]
XX
PR 08-JAN-1998; [illegible]
XX
PA (INME-) BIOMERIEUX INC.
XX
PI Gilmore RD, Johnson BJB;
XX
DR WPI; 1999-4[illegible]
XX
PT Recombinant FlaA/[illegible] protein antigen, useful [illegible] detection
PT of Borrelia burgdorferi
XX

SUMMARIES

	Description
AAX[illegible]	PCR primer [illegible]
AAX[illegible]	Nucleic acid seque
ABL[illegible]	Drosophila melanog
ABL[illegible]	Drosophila melanog
ABL[illegible]	Drosophila melanog
ABL[illegible]	Drosophila melanog
ABA[illegible]	Human sn-glycerol
AA[illegible]	Arabidopsis thalia
AA[illegible]	Arabidopsis thalia

```
PS   Example 2; Page 11; 46pp; English.
XX
CC   The sequence is a PCR primer denoted F1, used to create the F1
CC   construct, which consists of the entire coding sequence of P37
CC   (AAX89520). The P37 protein product can be used as an assay for the
CC   detection of Borrelia burgdorferi infection, the causative agent of Lyme
CC   disease. FlaA is an outer sheath protein of the periplasmic flagella of
CC   B.burgdoferi. FlaA is a prominent antigen in the early humoral immune
CC   response to Borrelia burgdorferi infection, and is significantly suitable
CC   for use in improved serologic tests for exposure to Lyme disease
CC   spirochetes. The detection of FlaA can augment the set of recombinant
CC   molecules that are recognized early in the course of disease and
CC   contribute to improved sensitivity of early testing for Lyme disease.
XX
SQ   Sequence 21 BP; 12 A; 1 C; 5 G; 3 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 20;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAAAGT 21
              |||||||||||||||||||||
Db          1 ATGAAAAGGAAAGCTAAAAGT 21


RESULT 2
AAX89520
ID   AAX89520 standard; DNA; 1655 BP.
XX
AC   AAX89520;
XX
DT   15-OCT-1999  (first entry)
XX
DE   Nucleic acid sequence encoding the FlaA gene product.
XX
KW   P37; FlaA; Lyme disease; antigen; assay; spirochete; ss.
XX
OS   Synthetic.
OS   Borrelia burgdorferi
XX
FH   Key             Location/Qualifiers
FT   CDS             473..1498
FT                   /*tag= a
FT                   /product= "P37/FlaA protein antigen"
FT   sig_peptide     473..538
FT                   /*tag= b
FT   mat_peptide     539..1495
FT                   /*tag= c
XX
PN   WO9935272-A1.
XX
PD   15-JUL-1999.
XX
PF   08-JAN-1999;   99WO-US00106.
XX
PR   08-JAN-1998;   98US-0004395.
XX
PA   (INMR ) BIOMERIEUX INC.
XX
PI   Gilmore RD,  Johnson BJB;
XX
DR   WPI; 1999-435430/36.
DR   P-PSDB; AAY28460.
XX
PT   Recombinant FlaA/P37 protein antigen, useful in improved detection
PT   of Borrelia burgdorferi
XX
PS   Disclosure; Page 28; 34pp; English.
XX
CC   The sequence encodes the recombinant protein antigen P37/FlaA. The
CC   P37 protein product can be used in an assay for the detection of Borrelia
CC   burgdorferi infection, the causative agent of Lyme disease. FlaA is
CC   an outer sheath protein of the periplasmic flagella of B. burgdorferi.
CC   FlaA is a prominent antigen in the early humoral immune response to
CC   Borrelia burgdorferi infection, and significantly suitable for use in
CC   improved serologic tests for exposure to Lyme disease spirochetes. The
CC   detection of FlaA can augment the set of recombinant molecules that are
CC   recognized early in the course of disease and contribute to improved
CC   sensitivity of early testing for Lyme disease.
XX
SQ   Sequence 1655 BP; 560 A; 182 C; 308 G; 605 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 20;  Length 1655;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAAAGT 21
              |||||||||||||||||||||
Db        473 ATGAAAAGGAAAGCTAAAAGT 493


RESULT 3
ABL12472
ID   ABL12472 standard; cDNA; 4133 BP.
XX
AC   ABL12472;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO [illegible]
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68369.
XX
PT   New isolated nucleic acid detection reagent for detecting [illegible]
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions
XX
PS   Claim 1; SEQ ID NO 21898; 2pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL[illegible]-ABL[illegible]), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB[illegible]-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4133 BP; 1625 A; 856 C; 1004 G; 1248 T; 0 other;

  Query Match             84.8%;  Score 17.8;  DB 22;  Length 4133;
  Best Local Similarity   90.5%;  Pred. No. [illegible];
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

[illegible]

```
DE   Drosophila melanogaster expressed [illegible] SEQ ID NO 47 [illegible]
XX
KW   Drosophila; developmental [illegible]
KW   pharmaceutical; [illegible]
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; [illegible] WO-US[illegible].
XX
PR   23-MAR-2000; [illegible] US-[illegible].
PR   [illegible]
XX
PA   (CELE-) CELERA GENOMICS CORP.
XX
PI   Venter JC, Adams M, [illegible]
XX
DR   WPI; 2001-[illegible]/[illegible].
DR   P-PSDB; ABB[illegible].
XX
PT   New isolated nucleic acid detection reagent [illegible]
PT   genes from Drosophila and [illegible] cell signalling and cell-cell
PT   interactions
XX
PS   Claim 1; SEQ ID NO 47; [illegible] Sequence Listing [illegible]
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting [illegible] genes from Drosophila. The invention is
CC   useful in developmental biology and in [illegible] cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL[illegible]-ABL[illegible]), expressed DNA
CC   sequences (ABL[illegible]-ABL[illegible]) and the encoded proteins
CC   (ABB57[illegible]-ABB[illegible]).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at [illegible]wipo.int/pat/published_pct_sequences.
XX
SQ   Sequence 121[illegible] BP; [illegible] A; [illegible] other;

  Query Match              84.8%;  Score [illegible];  DB [illegible];  Length 121[illegible];
  Best Local Similarity    [illegible]%;  Pred. No. [illegible];
  Matches [illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy          1 AT[illegible]AAAA[illegible]AAA[illegible]AAAA[illegible] [illegible]

Db       [illegible] AT[illegible]AAAA[illegible]AAAA[illegible]AAA[illegible] [illegible]


RESULT [illegible]
ABL124[illegible]
ID   ABL124[illegible] standard; cDNA; [illegible]
XX
AC   ABL124[illegible];
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed [illegible] SEQ ID NO [illegible]
XX
KW   Drosophila; developmental [illegible]
KW   pharmaceutical; [illegible]
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; [illegible] WO-US[illegible].
XX
```

```
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614159.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW.
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68377.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions
XX
PS   Claim 1; SEQ ID No 31922; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL15176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 13003 BP; 3368 A; 2522 C; 2993 G; 4120 T; 0 other;

  Query Match             84.8%;  Score 17.8;  DB 23;  Length 13003;
  Best Local Similarity   90.5%;  Pred. No. 2.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAAAGT 21
              |||| |||||||||| ||||
Db      12601 ATGAATAGGAAAGCTATAAGT 12621


RESULT 7
ABA08415
ID   ABA08415 standard; cDNA; 469 BP.
XX
AC   ABA08415;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human sn-glycerol-3-P-acyltransferase homologue cDNA, SEQ ID NO:191.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
```

```
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-477216/49.
DR   P-PSDB; ABB11171.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer
XX
PS   Claim 1; Page 407; 1966pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a DNA encoding a
CC   novel human polypeptide of the invention.
XX
SQ   Sequence 469 BP; 142 A; 98 C; 109 G; 120 T; 0 other;

  Query Match             82.9%;  Score 17.4;  DB 22;  Length 469;
  Best Local Similarity   94.7%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 TGAAAAGGAAAGCTAAAAG 20
              | |||| ||||||||||||
Db        266 TGAAAAAGAAAGCTAAAAG 284


RESULT 8
AAC49540/c
ID   AAC49540 standard; DNA; 629 BP.
XX
AC   AAC49540;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 41536.
XX
```

```
PR   01-SEP-1999;   99US-0151590.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156159.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             92.9%;  Score 17.1;  DB 21;  Length 626;
  Best Local Similarity   94.7%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 GAAAAGGAAAGCTAAAAGT 21
            |||||||| ||||||||||
Db      505 GAAAAGGAATGCTAAAAGT 487


RESULT 9
AAC50183/c
ID   AAC50183 standard; DNA; 627 BP.
XX
AC   AAC50183;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 63886.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
```

PR 22-SEP-1999; [illegible]
PR 23-SEP-1999; [illegible]
PR 23-SEP-1999; [illegible]
PR 24-SEP-1999; [illegible]
PR 24-SEP-1999; [illegible]
PR 04-OCT-1999; [illegible]
PR 05-OCT-1999; [illegible]
PR 06-OCT-1999; [illegible]
PR 07-OCT-1999; [illegible]
PR 08-OCT-1999; [illegible]
PR 12-OCT-1999; [illegible]
PR 13-OCT-1999; [illegible]
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; [illegible]
PR 14-OCT-1999; [illegible]
PR 14-OCT-1999; [illegible]
PR 14-OCT-1999; [illegible]
PR 14-OCT-1999; [illegible]
PR 14-OCT-1999; [illegible]
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; [illegible]
PR 21-OCT-1999; [illegible]
PR 21-OCT-1999; [illegible]
PR 21-OCT-1999; [illegible]
PR 21-OCT-1999; [illegible]
PR 21-OCT-1999; [illegible]
PR 22-OCT-1999; [illegible]
PR 22-OCT-1999; [illegible]
PR 22-OCT-1999; [illegible]
PR 25-OCT-1999; [illegible]
PR 25-OCT-1999; [illegible]
PR 25-OCT-1999; [illegible]
PR 26-OCT-1999; [illegible]
PR 26-OCT-1999; [illegible]
PR 26-OCT-1999; [illegible]
PR 28-OCT-1999; [illegible]
PR 28-OCT-1999; [illegible]
PR 28-OCT-1999; [illegible]
PR 29-OCT-1999; [illegible]

Query Match [illegible]; Score [illegible]; DB [illegible]; Length [illegible];
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Qy 3 GAAAAGAAAG[illegible]AAAA[illegible]
Db 507 GAAAAGAAAG[illegible]AAAAG[illegible] 48[illegible]

RESULT [illegible]
AA[illegible]
ID AA[illegible] standard; DNA; [illegible] BP.
XX
AC AA[illegible];
XX
DT [illegible] (first entry)
XX
DE Arabidopsis thaliana DNA fragment [illegible]
XX
KW Hybridisation assay; genetic mapping; gene expression; [illegible]
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; [illegible]
XX
OS Arabidopsis thaliana.
XX
PN EP[illegible] A2.
XX
PD [illegible]-SEP-200[illegible].
XX
PF [illegible]-FEB-200[illegible]
XX
PR 2[illegible]-FEB-1999; [illegible]
PR 05-MAR-1999; [illegible]

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130519.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132457.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134379.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 11-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142853.
PR 09-JUL-1999; 99US-0142926.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147234.
PR 04-AUG-1999; 99US-0147362.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148314.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149900.
PR 23-AUG-1999; 99US-0149933.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151088.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151945.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155134.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156590.
PR 04-OCT-1999; 99US-0157115.
PR 05-OCT-1999; 99US-0157753.

[illegible]

```
PT   neuroblastoma malignancy and susceptibility [illegible] tumor marker
PT   for anti cancer agents
XX
PS   Claim 1; Page [illegible]; [illegible]pp; Japanese.
XX
CC   The invention relates to novel genes (AA[illegible] AA[illegible]) expressed in
CC   human neuroblastoma. The nucleic acids are [illegible] as a probe or
CC   primer in diagnosing the prognosis of human neuroblastoma malignancy and
CC   susceptibility [illegible] tumor markers for anti cancer agents. The
CC   gene information for diagnosing prognosis is related [illegible] similar
CC   to that for N-myc and TrkA genes.
XX
SQ   Sequence [illegible]

  Query Match             [illegible];  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   [illegible];  Pred. No. [illegible];
  Matches   18;  Conservative    [illegible];  Mismatches   [illegible];  Indels   [illegible];  Gaps   [illegible];

Qy          2 [illegible]AAAA[illegible]AAA[illegible]AAAA[illegible]
                  [illegible]
Db        72[illegible] [illegible]AAAA[illegible]AAAN[illegible]AAA[illegible]


RESULT 12
ABK94821
ID   ABK94821 standard; DNA; [illegible] BP.
XX
AC   ABK94821;
XX
DT   29-AUG-2002  (first entry)
XX
DE   DNA encoding human glycerol-3-phosphate acyltransferase [illegible].
XX
KW   Fatty acid regulated gene; polyunsaturated fatty acid disorder;
KW   PUFA disorder; eczema; cardiovascular disorder; hypertriglyceridaemia;
KW   dyslipidaemia; atherosclerosis; coronary artery disease;
KW   cerebrovascular disease; peripheral vascular disease; inflammation;
KW   sinusitis; asthma; pancreatitis; osteoarthritis; rheumatoid arthritis;
KW   acne; body weight disorder; obesity; cachexia; anorexia;
KW   psychiatric disorder; cancer; cystic fibrosis; premenstrual syndrome;
KW   diabetes; diabetic complication; genetic polymorphism; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO[illegible] A[illegible].
XX
PD   2[illegible]-MAY-2002.
XX
PF   1[illegible]-NOV-20[illegible];  2[illegible]W[illegible]A[illegible].
XX
PR   17-NOV-2000; 2000US-248[illegible].
XX
PA   (XENO-) XENON GENETICS INC.
XX
PI   Winther MD, Goldberg YP, Knickle [illegible], Hayden M, A[illegible], [illegible] A;
PI   De Antueno RJ, Jenkins DK, Nwaka S;
XX
DR   WPI; 2002-[illegible]/[illegible]1.
DR   P-PSDB; AB[illegible].
XX
PT   Novel isolated polypeptide segment [illegible] fat regulated genes,
PT   useful for diagnosing the presence of or a predisposition for a
PT   disorder involving fatty acid regulated genes in a subject
XX
PS   Claim 1; Fig 28; 22[illegible]pp; English.
XX
CC   The invention describes an isolated polypeptide segment (I) whose genes
CC   are fat regulated. (I) or the polynucleotide encoding it (II) are useful
CC   for diagnosing the presence of or a predisposition for a disorder
CC   involving fatty acid regulated genes in a subject. A composition
CC   containing (I) or (II) is useful for treating a disorder involving fatty
CC   acid regulated genes, where the disorder is selected from a
CC   polyunsaturated fatty acid (PUFA) disorder, cardiovascular
```

```
CC   disorders (such as hypertriglyceridaemia, dyslipidaemia, atherosclerosis,
CC   coronary artery disease, cerebrovascular disease or peripheral vascular
CC   disease), inflammation (such as sinusitis, asthma, pancreatitis,
CC   osteoarthritis, rheumatoid arthritis or acne), body weight disorders
CC   (such as obesity, cachexia or anorexia), psychiatric disorders, cancer,
CC   cystic fibrosis, pre-menstrual syndrome, diabetes, and diabetic
CC   complications. (I) or (II) is useful as research agent and materials for
CC   discovery of treatments and diagnostics for a disease, particularly human
CC   disease. (II) is useful for constructing nucleotide probes and primers,
CC   for detecting genetic polymorphism, for detecting changes in the level of
CC   expression of (II), and as a diagnostic tool. This sequence encodes
CC   a protein regulated by fatty acids.
XX
SQ   Sequence 2487 BP; 706 A; 548 C; 570 G; 663 T; 0 other;

  Query Match             82.9%;  Score 17.4;  DB 24;  Length 2487;
  Best Local Similarity   94.7%;  Pred. No. 3.8e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 TGAAAAGGAAAGCTAAAAG 20
              |||||:|||||||||||||
Db       497 TGAAAAAGAAAGCTAAAAG 515


RESULT 13
ABK11094
ID   ABK11094 standard; cDNA; 3093 BP.
XX
AC   ABK11094;
XX
DT   18-JUN-2002  (first entry)
XX
DE   cDNA encoding human acyltransferase, ACTR 1.
XX
KW   Human; acyltransferase; ACTR 1; antidiabetic; antilipaemic; anorectic;
KW   antiarteriosclerotic; cerebroprotective; anorectic; ophthalmological;
KW   cardiant; metabolic disorder; energy homeostasis disorder; diabetes;
KW   hyperglycaemia; hypercholesterolaemia; hyperlipoproteinaemia; stroke;
KW   hypertriglyceridaemia; hyperlipidaemia; atherosclerosis; obesity;
KW   retinopathy; nephropathy; peripheral neuropathy; weight disorder;
KW   appetite regulation disorder; cachexia, anorexia, bulimia,
KW   cardiovascular disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             341..2827
FT                   /*tag=  a
FT                   /product= "Human acyltransferase ACTR 1"
FT                   /note= "The coding region is specifically claimed
FT                          in claim 1"
XX
PN   WO200216792-A2.
XX
PD   28-FEB-2002.
XX
PF   21-AUG-2001; 2001WO-US26239.
XX
PR   21-AUG-2000; 2000US-226509P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Kapeller-Libermann R,  Logan TJ;
XX
DR   WPI; 2002-280634/32.
DR   P-PSDB; AAU78393.
XX
PT   New acyltransferase 1, specifically glycerol-3-phosphate
PT   acyltransferase, nucleic acids and proteins, useful for treating or
PT   diagnosing e.g. metabolic disorders, weight disorders and appetite
PT   regulation disorders
XX
PS   Claim 1; Fig 1; 122pp; English.
XX
CC   The invention relates to acyltransferase (ACTR 1) nucleic acids and
CC   proteins. These are useful as modulating agents in regulating a variety
CC   of cellular processes (e.g. cellular metabolism, energy homeostasis
CC   and/or cellular proliferation, growth, differentiation and/or migration).
CC   The nucleic acids, proteins, protein homologues, fragments, antibodies,
CC   peptides, peptidomimetics and small molecules can be used in screening
CC   assays, predictive medicine (e.g. diagnostic or prognostic assays,
CC   monitoring clinical trials, and pharmacogenetics), and in methods of
CC   treatment (e.g. therapeutic or prophylactic). The ACTR 1 polypeptide or
CC   its fragments are useful as reagents or targets in assays for the
CC   treatment and/or diagnosis of ACTR 1 mediated or related disorders, which
CC   include metabolic disorders such as disorders of energy homeostasis e.g.
CC   diabetes, impaired glucose tolerance, insulin resistance, hyperglycaemia,
CC   hypercholesterolaemia, hyperlipoproteinaemia, hypertriglyceridaemia,
CC   and/or hyperlipidaemia; diabetic complications including atherosclerosis,
CC   stroke, retinopathy, nephropathy, and peripheral neuropathy; weight
CC   disorders and appetite regulation disorders, e.g. obesity, cachexia,
CC   anorexia and bulimia. Proteins may also be used to screen for naturally
CC   occurring ACTR 1 substrates, to screen for drugs or compounds which
CC   modulate ACTR 1 expression and to treat disorders characterised by
CC   insufficient or excessive production of ACTR 1 protein. The nucleic acids
CC   may be used as hybridisation probes or primers, to express ACTR 1
CC   protein, to detect ACTR 1 mRNA or a genetic alteration in an ACTR 1 gene,
CC   to modulate ACTR 1 activity, in chromosome mapping, tissue typing and
CC   forensic biology, and as surrogate markers (e.g. markers for precursors
CC   of disease states, for predisposition of disease states, of drug
CC   activity or of the pharmacogenomic profile of a subject). Modulators of
CC   ACTR 1 protein activity or nucleic acid expression may be used to treat
CC   a subject having a cardiovascular disorder or a triglyceride metabolism
CC   disorder characterised by aberrant ACTR-1 protein activity or nucleic
CC   acid expression. Anti-ACTR-1 antibodies can be used to detect and
CC   isolate ACTR 1 proteins, regulate the bioavailability of ACTR 1
CC   proteins, and modulate ACTR 1 activity. The present sequence
CC   represents the coding sequence of human ACTR 1.
XX
SQ   Sequence 3093 BP; 826 A; 683 C; 698 G; 794 T; 2 other;

  Query Match             82.9%;  Score 17.4;  DB 24;  Length 3093;
  Best Local Similarity   94.7%;  Pred. No. 3.8e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 TGAAAAGGAAAGCTAAAAG 20
              |||||:|||||||||||||
Db        837 TGAAAAAGAAAGCTAAAAG 855


RESULT 14
AAD35221
ID   AAD35221 standard; cDNA; 3210 BP.
XX
AC   AAD35221;
XX
DT   25-JUL-2002  (first entry)
XX
DE   Human TRNFR 6 cDNA.
XX
KW   Human; transferase; developmental disorder; cell proliferative disorder;
KW   TRNFR-6; neurological disorder; autoimmune disorder; parasitic infection;
KW   inflammatory disorder; endocrine; antiparasitic; immunosuppressive;
KW   cytostatic; neurological; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             358..1824
FT                   /*tag=  a
FT                   /product= "Human TRNFR 6 protein"
XX
PN   WO200226950-A2.
XX
PD   04-APR-2002.
XX
```

PI Schlegel R, Endege W, Monahan [illegible]
XX
DR WPI; 2001-66279[illegible]
XX
PT Novel isolated nucleic acid molecule associated with [illegible] state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer [illegible] of prostate cancer
XX
PS Claim 1; Page 498 [illegible]; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables [illegible] (ABV[illegible]-ABV[illegible]) of the
CC specification or its complement. (I) is useful for
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound for inhibiting prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient
CC (f) assessing the prostate cell carcinogenic potential of a compound
CC (g) determining whether prostate cancer has metastasized in a patient
CC (h) assessing the aggressiveness of [illegible] of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 4[illegible] BP; [illegible]

Query Match [illegible]; Score [illegible]; DB [illegible]; Length 4[illegible];
Best Local Similarity [illegible]4.7%; Pred. No. [illegible];
Matches 18; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

```
Qy          2 TGAAAAGGAAAGTAAAA[illegible]
              ||||||  ||||||||
Db        [illegible]7 TGAAAAAGAAAGTAAAA[illegible]
```

Search completed: November [illegible]
Job time : [illegible] secs

[illegible]

(without alignments)

[illegible] updates/se[illegible]

[illegible] being printed,

SUMMARY [illegible]

Description

[illegible]	Sequence 5, Appli
[illegible]	Sequence 5, Appli
[illegible]	Sequence 5, Appli
[illegible]	Sequence 5, Appli
[illegible]	Sequence 6, Appli
[illegible]	Sequence 6, Appli
[illegible]	Sequence 4, Appli
[illegible]	Sequence 8, Appli
[illegible]	Sequence 5, Appli
[illegible]	Sequence 5, Appli
[illegible]	Sequence 7, Appli
[illegible]	Sequence 5, Appli
[illegible]	Sequence 9, Appli
[illegible]	Sequence 9, Appli
[illegible]	Sequence 9, Appli
[illegible]	Sequence 6, Appli
[illegible]	Sequence 4, Appli
[illegible]	Sequence 4, Appli
[illegible]	Sequence 4, Appli
[illegible]	Sequence 4, Appli
[illegible]	Sequence 1, Appli
[illegible]	Sequence 1, Appli
[illegible]	Sequence 15, Appl
[illegible]	Sequence 1, Appli
[illegible]	Sequence 1, Appli
[illegible]	Sequence 5, Appli
[illegible]	Sequence 55, Appl

28	15.8	79.2	[illegible]	[illegible]	Sequence [illegible], Appl
29	15.8	79.2	1554	[illegible]	Sequence [illegible], Appl
30	15.8	79.2	17[illegible]	[illegible]	Sequence 1, Appl
31	15.8	79.2	[illegible]	[illegible]	Sequence 1, Appl
32	15.8	79.2	17[illegible]	[illegible]	Sequence 1, Appl
33	15.8	79.2	17[illegible]	[illegible]	Sequence 1, Appl
34	15.8	79.2	17[illegible]	[illegible]	Sequence 4, Appl
35	15.4	75.3	44[illegible]	4 US [illegible]	Sequence [illegible], Appl
36	15.4	75.3	1214[illegible]	[illegible]	Sequence 1[illegible], Appl
37	15.4	75.3	1214[illegible]	[illegible]	Sequence 1[illegible], Appl
38	15.2	72.4	[illegible]	[illegible]	Sequence 1, Appl
39	15.2	72.4	[illegible]	[illegible]	Sequence 1, Appl
40	15.2	72.4	[illegible]	[illegible]	Sequence 1, Appl
41	15.2	72.4	[illegible]	[illegible]	Sequence 1, Appl
42	15.2	72.4	1902	[illegible]	Sequence 1, Appl
43	15.2	72.4	[illegible]	[illegible]	Sequence 1, Appl
44	15.2	72.4	[illegible]	[illegible]	Sequence 1, Appl
45	15.2	72.4	[illegible]	[illegible]	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-08-404-[illegible]A-[illegible]
; Sequence [illegible], Application US/08404[illegible]A
; Patent No. 5661[illegible]
;  GENERAL INFORMATION:
;    APPLICANT: [illegible]
;    APPLICANT: Roessler, [illegible]
;    APPLICANT: [illegible]
;    TITLE OF INVENTION: METHOD [illegible] TRANSFORMATION [illegible] MATERIAL
;    TITLE OF INVENTION: THEREFOR AND [illegible] THEREBY
;    NUMBER OF SEQUENCES: [illegible]
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: National Renewable Energy Laboratory
;      STREET: 1617 Cole Boulevard
;      CITY: Golden
;      STATE: [illegible]
;      COUNTRY: U.S.A.
;      ZIP: [illegible]
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM [illegible]
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: ASCII([illegible])text (*.*)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/404,[illegible]A
;      FILING DATE: [illegible]
;      CLASSIFICATION: 4[illegible]
;    ATTORNEY/AGENT INFORMATION:
;      NAME: O'Connor, Edna M.
;      REGISTRATION NUMBER: [illegible]
;      REFERENCE/DOCKET NUMBER: [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: [illegible]/275-44[illegible]
;      TELEFAX: [illegible]/275-[illegible]44
;  INFORMATION FOR SEQ ID NO: [illegible]:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: [illegible]4 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: [illegible]
;      TOPOLOGY: [illegible]
;    MOLECULE TYPE: DNA (genomic)
;    FEATURE:
;      NAME/KEY: [illegible]
;      LOCATION: 1..[illegible]4
US-08-404-[illegible]A-[illegible]

  Query Match [illegible]%; Score [illegible]; DB [illegible]; Length [illegible]4;
  Best Local Similarity [illegible]%; Pred. No. [illegible];
  Matches [illegible]8; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];
```

```
Qy       1 atgaaaaggaaagctaaaagt 21
Db     100 ATGAAAAGCAATCCTAAAAGT 80

RESULT  2
US-08-080-255-3/c
; Sequence 3, Application US/08080255
; Patent No. 5487978
;  GENERAL INFORMATION:
;    APPLICANT: Rowley, Janet D.
;    APPLICANT: Diaz, Manuel O.
;    TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
;    TITLE OF INVENTION: TRANSLOCATIONS
;    NUMBER OF SEQUENCES: 8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Arnold, White & Durkee
;      STREET: P. O. Box 4433
;      CITY: Houston
;      STATE: Texas
;      COUNTRY: USA
;      ZIP: 77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/080,255
;      FILING DATE: 19930617
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Parker, David L.
;      REGISTRATION NUMBER: 32,165
;      REFERENCE/DOCKET NUMBER: ARCD:072/TAE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (512) 320-7200
;      TELEFAX: (512) 474-7577
;  INFORMATION FOR SEQ ID NO: 3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1420 base pairs
;      TYPE: NUCLEIC ACID
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
US-08-080-255-3

  Query Match             77.1%;  Score 16.2;  DB 1;  Length 1420;
  Best Local Similarity   85.7%;  Pred. No. 98;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 atgaaaaggaaagctaaaagt 21
Db      70 ATGAGAAGGAAAAGTAAAAGT 50

RESULT  3
US-08-465-713-3/c
; Sequence 3, Application US/08465713
; Patent No. 6121419
;  GENERAL INFORMATION:
;    APPLICANT: Rowley, Janet D.
;    APPLICANT: Diaz, Manuel O.
;    TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
;    TITLE OF INVENTION: TRANSLOCATIONS
;    NUMBER OF SEQUENCES: 8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Arnold, White & Durkee
;      STREET: P. O. Box 4433
;      CITY: Houston
;      STATE: Texas
;      COUNTRY: USA
;      ZIP: 77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/465,713
;      FILING DATE: 06-JUN-1995
;      CLASSIFICATION: 536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US/08/080,255
;      FILING DATE: 17 JUNE 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Parker, David L.
;      REGISTRATION NUMBER: 32,165
;      REFERENCE/DOCKET NUMBER: ARCD:072/TAE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (512) 320-7200
;      TELEFAX: (512) 474-7577
;  INFORMATION FOR SEQ ID NO: 3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1420 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
US-08-465-713-3

  Query Match             77.1%;  Score 16.2;  DB 3;  Length 1420;
  Best Local Similarity   85.7%;  Pred. No. 98;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 atgaaaaggaaagctaaaagt 21
Db      70 ATGAGAAGGAAAAGTAAAAGT 50

RESULT  4
PCT-US94-06857-3/c
; Sequence 3, Application PC/TUS9406857
;  GENERAL INFORMATION:
;    APPLICANT: Board of Regents
;    APPLICANT: The University of Texas System
;    TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
;    NUMBER OF SEQUENCES: 8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Arnold, White & Durkee
;      STREET: P. O. Box 4433
;      CITY: Houston
;      STATE: Texas
;      COUNTRY: USA
;      ZIP: 77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: PCT/US94/06857
;      FILING DATE: 19940617
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 07/900,699
;      FILING DATE: 17/06/92
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Parker, David L.
;      REGISTRATION NUMBER: 32,165
```

[illegible]

; LOCATION: 1..4[illegible]
; NAME/KEY: A[illegible]
; NAME/KEY: [illegible]
; LOCATION: 1,44[illegible]
; NAME/KEY: [illegible]
; LOCATION: 1,6[illegible]..18[illegible]
; OTHER INFORMATION: [illegible] A[illegible]
US-08-404-7[illegible]2A [illegible]

Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy 1 [illegible] 21

Db [illegible] A[illegible]AAAA[illegible]AAAA[illegible]

RESULT [illegible]
US-09-[illegible]742-[illegible]/[illegible]
; Sequence 6, Application US/[illegible]274.
; Patent No. 62[illegible]
; GENERAL INFORMATION:
; APPLICANT: [illegible] M.[illegible]
; APPLICANT: [illegible]
; APPLICANT: [illegible]
; TITLE OF INVENTION: [illegible]
; TITLE OF INVENTION: Receptors
; NUMBER OF SEQUENCES: [illegible]
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: [illegible] Weinberg[illegible]
; STREET: [illegible] S. Matlack Street [illegible]
; STREET: P.O. Box [illegible]
; CITY: West Chester
; STATE: PA
; COUNTRY: USA
; ZIP: 193[illegible]
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #[illegible] Version #[illegible]
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/[illegible]
; FILING DATE: [illegible] FEB [illegible]
; CLASSIFICATION: [illegible]
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinberg[illegible]
; REGISTRATION NUMBER: [illegible]
; REFERENCE/DOCKET NUMBER: [illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (6[illegible]) 4[illegible]
; TELEFAX: (610) 4[illegible]
; INFORMATION FOR SEQ ID NO: [illegible]
; SEQUENCE CHARACTERISTICS:
; LENGTH: [illegible] base pairs
; TYPE: nucleic acid
; STRANDEDNESS: [illegible]
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-[illegible]742-6

Query Match [illegible]; Score [illegible]; DB 4[illegible]; Length 2[illegible]
Best Local Similarity [illegible]; Pred. No. [illegible]
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy 1 [illegible] 21

Db 1975 ATGAAAAGGA[illegible]AAAAG[illegible] 1[illegible]

```
RESULT  7
US-08-404-732A-4/c
; Sequence 4, Application US/08404732A
; Patent No. 5661017
;  GENERAL INFORMATION:
;    APPLICANT: Dunahay, Terry G.
;    APPLICANT: Roessler, Paul
;    APPLICANT: Jarvis, Eric
;    TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE, MATERIALS
;    TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY
;    NUMBER OF SEQUENCES: 11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: National Renewable Energy Laboratory
;      STREET: 1617 Cole Boulevard
;      CITY: Golden
;      STATE: CO
;      COUNTRY: U.S.A.
;      ZIP: 80401-3393
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: ASCII(DOS)text (*.*)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/404,732A
;      FILING DATE: 15-MAR-1995
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: O'Connor, Edna M.
;      REGISTRATION NUMBER: 29252
;      REFERENCE/DOCKET NUMBER: 94-31
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 303/275-4404
;      TELEFAX: 303/275-4400
;  INFORMATION FOR SEQ ID NO: 4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 2120 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
;    FEATURE:
;    FEATURE: 1.
;      NAME/KEY: 5'UTR
;      LOCATION: 1..445
;      OTHER INFORMATION: /label= ACCase
;      NAME/KEY: CDS
;      LOCATION: 446..1240
;      OTHER INFORMATION: /label= NPTII
;      NAME/KEY: Linker Sequence
;      LOCATION: 1241..1248
;      NAME/KEY: ACCase 3' coding region (in non-translatable reading
;      NAME/KEY: frame)
;      LOCATION: 1249..1526
;      NAME/KEY: 3'UTR
;      LOCATION: 1527..2120
;      OTHER INFORMATION: /label= ACCase
US-08-404-732A-4

  Query Match             77.1%;  Score 16.2;  DB 1;  Length 2120;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 atgaaaaagaaagctaaaagt 21
              |||||||| ||  |||||||
Db       1626 ATGAAAAGCAATCCTAAAAGT 1606

RESULT  8
US-08-404-732A-8/c
; Sequence 8, Application US/08404732A
; Patent No. 5661017
;  GENERAL INFORMATION:
;    APPLICANT: Dunahay, Terry G.
;    APPLICANT: Roessler, Paul
;    APPLICANT: Jarvis, Eric
;    TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE, MATERIALS
;    TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY
;    NUMBER OF SEQUENCES: 11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: National Renewable Energy Laboratory
;      STREET: 1617 Cole Boulevard
;      CITY: Golden
;      STATE: CO
;      COUNTRY: U.S.A.
;      ZIP: 80401-3393
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: ASCII(DOS)text (*.*)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/404,732A
;      FILING DATE: 15-MAR-1995
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: O'Connor, Edna M.
;      REGISTRATION NUMBER: 29252
;      REFERENCE/DOCKET NUMBER: 94-31
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 303/275-4404
;      TELEFAX: 303/275-4400
;  INFORMATION FOR SEQ ID NO: 8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 2224 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
;    FEATURE:
;    FEATURE: 1.
;      NAME/KEY: 5'UTR
;      LOCATION: 1..816
;      OTHER INFORMATION: /label= ACCase
;      NAME/KEY: CDS
;      LOCATION: 817..1611
;      OTHER INFORMATION: /label= NPTII
;      NAME/KEY: Linker Sequence
;      LOCATION: 1612..1614
;      NAME/KEY: ACCase 3' coding region (in non-translatable reading
;      NAME/KEY: frame)
;      LOCATION: 1615..1630
;      NAME/KEY: 3'UTR
;      LOCATION: 1631..2224
;      OTHER INFORMATION: /label= ACCase
US-08-404-732A-8

  Query Match             77.1%;  Score 16.2;  DB 1;  Length 2224;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 atgaaaaagaaagctaaaagt 21
              |||||||| ||  |||||||
Db       1730 ATGAAAAGCAATCCTAAAAGT 1710

RESULT  9
US-08-966-008-5
; Sequence 5, Application US/08966008
; Patent No. 5965412
;  GENERAL INFORMATION:
;    APPLICANT: Tomoyuki NISHIMOTO
;    APPLICANT: Michio KUBOTA
;    APPLICANT: Hiroto CHAEN
```

```
[illegible] PREPARATION AND USES

[illegible] NEIMARK
[illegible] N.W. [illegible]

[illegible] PREPARATION AND USES

[illegible] NEIMARK
[illegible] N.W. [illegible]
```

```
;        SOFTWARE: PatentIn Release #[illegible], Version #[illegible]
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/[illegible]
;        FILING DATE:
;        CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: [illegible]
;        FILING DATE:
;      PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: [illegible]
;        FILING DATE: [illegible] MAR [illegible]
;      ATTORNEY/AGENT INFORMATION:
;        NAME: BROWDY, Roger L.
;        REGISTRATION NUMBER: [illegible]
;        REFERENCE/DOCKET NUMBER:
;      TELECOMMUNICATION INFORMATION:
;        TELEPHONE: [illegible]
;        TELEFAX: [illegible]
;        TELEX: [illegible]
;    INFORMATION FOR SEQ ID NO: [illegible]
;      SEQUENCE CHARACTERISTICS:
;        LENGTH: [illegible] base pairs
;        TYPE: nucleic acid
;        STRANDEDNESS: [illegible]
;        TOPOLOGY: linear
US-09-188-4[illegible]

  Query Match [illegible]
  Best Local Similarity [illegible]
  Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy        1 [illegible] 21

Db     2274 A[illegible]GAAAA[illegible]AAA[illegible]A[illegible]AAAA[illegible] 22[illegible]4


RESULT 11
US-09-188-4[illegible]
; Sequence 5, Application US/09188[illegible]4
; Patent No. [illegible]
;   GENERAL INFORMATION:
;      APPLICANT: [illegible]
;      APPLICANT: [illegible] KUBOTA
;      APPLICANT: [illegible] CHAEN
;      APPLICANT: Toshio MIYAKE
;      TITLE OF INVENTION: [illegible] PREPARATION AND USES
;      NUMBER OF SEQUENCES: [illegible]
;      CORRESPONDENCE ADDRESS:
;        ADDRESSEE: BROWDY AND NEIMARK
;        STREET: 41[illegible] Seventh Street, N.W., Suite [illegible]
;        CITY: Washington
;        STATE: D.C.
;        COUNTRY: USA
;        ZIP: [illegible]
;      COMPUTER READABLE FORM:
;        MEDIUM TYPE: Floppy disk
;        COMPUTER: IBM PC compatible
;        OPERATING SYSTEM: PC-DOS/MS-DOS
;        SOFTWARE: PatentIn Release #[illegible], Version #[illegible]
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/[illegible]
;        FILING DATE:
;        CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: [illegible]
;        FILING DATE:
;      PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: [illegible]
;        FILING DATE: [illegible] MAR [illegible]
;      ATTORNEY/AGENT INFORMATION:
;        NAME: BROWDY, Roger L.
;        REGISTRATION NUMBER: [illegible]
```

```
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 202 628 5197
;       TELEFAX: 202 737 3528
;       TELEX: 248633
;   INFORMATION FOR SEQ ID NO: 5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 2325 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
US-09-288-401-5

  Query Match             77.1%;  Score 16.2;  DB 3;  Length 2325;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 atgaaaaggaaaatataaaagt 21
              | |.|.||||||.| ||||||
Db       2274 AGGAAAAGGAAAATATAAAAGT 2294

RESULT 12
US-09-281-259-5
; Sequence 5, Application US/09281259
; Patent No. 6204377
;  GENERAL INFORMATION:
;     APPLICANT: Tomoyuki NISHIMOTO
;     APPLICANT: Michio KUBOTA
;     APPLICANT: Hiroto CHAEN
;     APPLICANT: Toshio MIYAKE
;     TITLE OF INVENTION: RECOMBINANT THERMOSTABLE, ITS PREPARATION AND USES
;     NUMBER OF SEQUENCES: 10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: BROWDY AND NEIMARK
;       STREET: 419 Seventh Street, N.W., Suite 300
;       CITY: Washington
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 20004
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: Patent In Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/281,259
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/966,388
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: JP 61,716/97
;       FILING DATE: 3-MAR-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME: BROWDY, Roger L.
;       REGISTRATION NUMBER: 25,618
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 202 628 5197
;       TELEFAX: 202 737 3528
;       TELEX: 248633
;   INFORMATION FOR SEQ ID NO: 5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 2325 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
US-09-281-259-5

  Query Match             77.1%;  Score 16.2;  DB 4;  Length 2325;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 atgaaaaggaaaatataaaagt 21
              | |.|.||||||.| ||||||
Db       2274 AGGAAAAGGAAAATATAAAAGT 2294

RESULT 13
US-08-966-388-9
; Sequence 9, Application US/08966388
; Patent No. 5965412
;  GENERAL INFORMATION:
;     APPLICANT: Tomoyuki NISHIMOTO
;     APPLICANT: Michio KUBOTA
;     APPLICANT: Hiroto CHAEN
;     APPLICANT: Toshio MIYAKE
;     TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES
;     NUMBER OF SEQUENCES: 10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: BROWDY AND NEIMARK
;       STREET: 419 Seventh Street, N.W., Suite 300
;       CITY: Washington
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 20004
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: Patent In Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/966,388
;       FILING DATE:
;       CLASSIFICATION: 435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: JP 311,257/1996
;       FILING DATE: 8 NOV 1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: JP 61,716/97
;       FILING DATE: 3 MAR 1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME: BROWDY, Roger L.
;       REGISTRATION NUMBER: 25,618
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 202 628 5197
;       TELEFAX: 202-737-3528
;       TELEX: 248633
;   INFORMATION FOR SEQ ID NO: 9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 3956 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
;     MOLECULE TYPE: Genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM: Thermoanaerobium brockii
;       STRAIN: ATCC 35047
;     FEATURE:
;     FEATURE: 1 100 5' UTR
;     FEATURE: E
;     FEATURE: 101 2425 mat peptide
;     FEATURE: S
;     FEATURE: 2426 3956 3'UTR
;     FEATURE: E
US-08-966-388-9

  Query Match             77.1%;  Score 16.2;  DB 2;  Length 3956;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Db     2374 A[illegible]GAAAA[illegible]GAAATATAAAA[illegible] [illegible]

RESULT 15
US-09-188-404-[illegible]
; Sequence [illegible], Application US/09188404[illegible]
; Patent No. 614[illegible]
; GENERAL INFORMATION:
;    APPLICANT:  Tomoyuki NISHIMOTO
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  [illegible]
;    NUMBER OF SEQUENCES:  [illegible]
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  41[illegible] Seventh Street, N.W. [illegible]
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  2[illegible]4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.[illegible]
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/188,4[illegible]
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  [illegible]
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  [illegible]
;      FILING DATE:  [illegible] MAR [illegible]
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  [illegible]
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  [illegible]
;      TELEFAX:  [illegible]
;      TELEX:  24[illegible]
; INFORMATION FOR SEQ ID NO: [illegible]:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  [illegible] base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  [illegible]
;      STRAIN:  ATCC [illegible]
;    FEATURE:
;    FEATURE:  [illegible] CDS
;    FEATURE:  [illegible]
;    FEATURE:  [illegible] mat peptide
;    FEATURE:  S
;    FEATURE:  [illegible]
;    FEATURE:  [illegible]
US-09-188-404-[illegible]

  Query Match             [illegible]
  Best Local Similarity   [illegible]
  Matches   18;  Conservative    [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy         1 [illegible] [illegible]
             [illegible]
Db      2374 AGGAAAA[illegible]GAAATATAAAA[illegible] [illegible]
```

Search completed: March 2[illegible], 2002, [illegible]:39:39
Job time: [illegible] sec

[illegible]

SUMMARY

Result No.	Score	Query Match	Length	DB	ID	Description
20	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
21	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
22	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
23	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
24	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]-4414	Sequence 4414, Ap
25	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
26	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
27	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
28	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
29	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
30	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
31	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
32	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
33	16.2	[illegible]	4465	[illegible]	US-[illegible]	Sequence [illegible], Ap
34	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
35	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
36	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
37	16.2	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
38	16	26.2	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
39	15.8	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], A
40	15.8	25.2	403	[illegible]	US-[illegible]	Sequence [illegible], App
41	15.8	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], Ap
42	15.8	25.2	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
43	15.8	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
44	15.8	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App
45	15.8	[illegible]	[illegible]	[illegible]	US-[illegible]	Sequence [illegible], App

ALIGNMENTS

RESULT 1
US-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. US20[illegible]A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES [illegible]
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: [illegible]
; CURRENT APPLICATION NUMBER: US/[illegible]
; CURRENT FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: US [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: US [illegible]
; PRIOR FILING DATE: 2000-[illegible]
; PRIOR APPLICATION NUMBER: US [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: GB [illegible]
; PRIOR FILING DATE: 2000-[illegible]
; PRIOR APPLICATION NUMBER: US [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]
; PRIOR APPLICATION NUMBER: PCT/US01/[illegible]
; PRIOR FILING DATE: 2001-[illegible]

```
;   PRIOR APPLICATION NUMBER: [illegible]
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: US 60/234,687
;   PRIOR FILING DATE: 2000-09-21
;   PRIOR APPLICATION NUMBER: US 09/608,408
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: US 09/774,203
;   PRIOR FILING DATE: 2001-01-29
;   NUMBER OF SEQ ID NOS: 49117
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10375
;    LENGTH: 353
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;     OTHER INFORMATION: MAP TO AC[illegible]
;     OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL 1.1
;     OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL 1.1
;     OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL 0.92
;     OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL 1.1
;     OTHER INFORMATION: EXPRESSED IN H1474, SIGNAL 0.97
;     OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL 1.1
US-09-864-761-10375

  Query Match             [illegible]%;  Score 16.8;  DB 10;  Length 353;
  Best Local Similarity   90.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TGAAAAGGAAAGCTAAAAGT 21
              |||||||   |||||| |
Db        145 TGAAAAGGTAGGCTAAAAGT 164


RESULT 2
US-09-864-761-10045
; Sequence 10045, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE: 2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/[illegible],812
;  PRIOR FILING DATE: 2000-[illegible]
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-[illegible]
;  PRIOR APPLICATION NUMBER: GB [illegible]
;  PRIOR FILING DATE: 2000-[illegible]
;  PRIOR APPLICATION NUMBER: US 60/236,[illegible]
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00660
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10045
;    LENGTH: 476
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;     OTHER INFORMATION: MAP TO AC010062.1
;     OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL [illegible]
;     OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL [illegible]
;     OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL [illegible]
;     OTHER INFORMATION: EXPRESSED IN H1474, SIGNAL [illegible]
;     OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL 0.96
;     OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL 1.6
;     OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL [illegible]
;     OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL [illegible]
US-09-864-761-10045

  Query Match             80.0%;  Score 16.8;  DB [illegible];  Length 476;
  Best Local Similarity   90.0%;  Pred. No. [illegible];
  Matches   18;  Conservative    [illegible];  Mismatches    [illegible];  Indels    [illegible];  Gaps    [illegible];

Qy          2 TGAAAAGGAAAGCTAAAAGT 21
              |||||||   |||||| |
Db        159 TGAAAAGGTAGGCTAAAAGT 178


RESULT 3
US-09-954-456-2178/c
; Sequence 2178, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using [illegible]
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 609290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE: 2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,61[illegible]
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,92[illegible]
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,1[illegible]4
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/[illegible]
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,6[illegible]8
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,72[illegible]
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,84[illegible]
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,86[illegible]
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2178
;    LENGTH: 5421
;    TYPE: DNA
```

```
; Patent No. US[illegible]
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R
;  APPLICANT: [illegible]
;  APPLICANT: Thompson, Michael [illegible]
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION: Plants
;  FILE REFERENCE: [illegible]
;  CURRENT APPLICATION NUMBER: US/[illegible]
;  CURRENT FILING DATE: [illegible]
;  PRIOR APPLICATION NUMBER: [illegible]
;  PRIOR FILING DATE: [illegible]
;  NUMBER OF SEQ ID NOS: [illegible]
; SEQ ID NO [illegible]
;   LENGTH: 264
;   TYPE: DNA
;   ORGANISM: [illegible]
;   OTHER INFORMATION: [illegible]
US-09-878-[illegible]

  Query Match            [illegible];  Score [illegible];  DB [illegible];  Length 264;
  Best Local Similarity   94.4%;  Pred. No. [illegible];
  Matches   17;  Conservative    [illegible];  Mismatches    [illegible];  Indels    [illegible];  Gaps    [illegible];

Qy          1 ATGAAAATGAAATTAAA 1[illegible]
              [illegible]
Db        [illegible] ATGAAAATGAAATTGAA [illegible]

RESULT [illegible]
US-09-867-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. US[illegible]
; GENERAL INFORMATION:
;  APPLICANT: Afar, [illegible] A
;  APPLICANT: Jones, Robert
;  APPLICANT: Harlocker, Susan L
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 21[illegible]
;  CURRENT APPLICATION NUMBER: US/[illegible]
;  CURRENT FILING DATE: 2001-[illegible]
;  NUMBER OF SEQ ID NOS: [illegible]
;  SOFTWARE: FastSEQ for Windows Version 4.[illegible]
; SEQ ID NO [illegible]
;   LENGTH: 464
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-867-[illegible]

  Query Match            [illegible];  Score [illegible];  DB [illegible];  Length 464;
  Best Local Similarity   94.4%;  Pred. No. [illegible];
  Matches   17;  Conservative    [illegible];  Mismatches    [illegible];  Indels    [illegible];  Gaps    [illegible];

Qy          1 ATGAAAATGAAATTAAA 1[illegible]
              [illegible]
Db        [illegible] ATGAAAATGAAATTAAA [illegible]

RESULT [illegible]
US-09-867-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. US20020132237A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, [illegible] A.
;  APPLICANT: Jones, Robert
;  APPLICANT: Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 21[illegible]
;  CURRENT APPLICATION NUMBER: US/[illegible]
;  CURRENT FILING DATE: 2001-[illegible]
;  NUMBER OF SEQ ID NOS: [illegible]
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6057
;   LENGTH: 485
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-867-701-6057

  Query Match             78.1%;  Score 16.4;  DB 10;  Length 485;
  Best Local Similarity   94.4%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAA 18
              ||||||| ||||||||||
Db        230 ATGAAAAGCAAAGCTAAA 247

RESULT 9
US-09-777-564-851
; Sequence 851, Application US/09777564
; Patent No. US20020225591A1
; GENERAL INFORMATION:
;  APPLICANT: Algate, Paul A.
;  APPLICANT: Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.493
;  CURRENT APPLICATION NUMBER: US/09/777,564
;  CURRENT FILING DATE: 2001-02-05
;  NUMBER OF SEQ ID NOS: 1730
;  SOFTWARE: FastSEQ for Window Version 4.0
; SEQ ID NO 851
;   LENGTH: 525
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(525)
;   OTHER INFORMATION: n = A,T,C or G
US-09-777-564-851

  Query Match             78.1%;  Score 16.4;  DB 10;  Length 525;
  Best Local Similarity   94.4%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAA 18
              ||||||| ||||||||||
Db         25 ATGAAAAGCAAAGCTAAA 42

RESULT 10
US-10-046-935-185/c
; Sequence 185, Application US/10046935
; Patent No. US20020150911A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT: Secrist, Heather
;  APPLICANT: Wang, Aijun
;  APPLICANT: Stolk, John A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.52701
;  CURRENT APPLICATION NUMBER: US/10/046,935
;  CURRENT FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 2239
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 185
;   LENGTH: 557
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 521
;   OTHER INFORMATION: n = A,T,C or G
US-10-046-935-185

  Query Match             78.1%;  Score 16.4;  DB 10;  Length 557;
  Best Local Similarity   94.4%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAA 18
              ||||||| ||||||||||
Db        236 ATGAAAAGCAAAGCTAAA 219

RESULT 11
US-09-604-287A-63/c
; Sequence 63, Application US/09604287A
; Patent No. US20020064872A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT: Dillon, Davin C.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: Xu, Jiangchun
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT: Hepler, William T.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
;  FILE REFERENCE: 210121.47007
;  CURRENT APPLICATION NUMBER: US/09/604,287A
;  CURRENT FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 489
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63
;   LENGTH: 683
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-604-287A-63

  Query Match             78.1%;  Score 16.4;  DB 10;  Length 683;
  Best Local Similarity   94.4%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAA 18
              ||||||| ||||||||||
Db        236 ATGAAAAGCAAAGCTAAA 219

RESULT 12
US-09-820-989A-3/c
; Sequence 3, Application US/09820989A
; Patent No. US20020072503A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT: Stolk, John A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE: 210121.509
;  CURRENT APPLICATION NUMBER: US/09/820,989A
;  CURRENT FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 3
;   LENGTH: 683
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-820-989A-3

  Query Match             78.1%;  Score 16.4;  DB 10;  Length 683;
  Best Local Similarity   94.4%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAA 18
              ||||||| ||||||||||
Db        236 ATGAAAAGCAAAGCTAAA 219
```

[illegible] AND METHODS FOR THEIR USE

[illegible] ANT [illegible] THERAPY

```
; [illegible] N [illegible] A.
; GENERAL INFORMATION:
;  APPLICANT: [illegible]
;  APPLICANT: [illegible]
;  APPLICANT: [illegible] Andrew
;  APPLICANT: [illegible]
;  TITLE OF INVENTION: NOVEL GENES, [illegible] AND METHODS [illegible]
;  TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION: OF HUMAN CANCERS
;  FILE REFERENCE: MRI-016B
;  CURRENT APPLICATION NUMBER: [illegible]
;  CURRENT FILING DATE: [illegible]
;  PRIOR APPLICATION NUMBER: [illegible]
;  PRIOR FILING DATE: [illegible]-4-14
;  NUMBER OF SEQ ID NOS: [illegible]
;  SOFTWARE: FastSEQ for Windows Version [illegible]
; SEQ ID NO [illegible]
;   LENGTH: [illegible]
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...([illegible])
;   OTHER INFORMATION: [illegible]
US-09-834-[illegible]

  Query Match             [illegible];  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   [illegible];  Pred. No. [illegible];
  Matches  [illegible];  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy          [illegible] A[illegible]AAAA[illegible]AAA[illegible]AAA [illegible]

Db      [illegible] A[illegible]AAAA[illegible]AAA[illegible]AAA [illegible]

Search completed: November [illegible]
Job time : [illegible] secs
```

[illegible]

ALIGNMENTS

```
RESULT      1
INS INNO
LOCUS       INS INNO   [illegible]
DEFINITION  Anopheles gambiae [illegible]
            from strain [illegible] of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL1[illegible]442
VERSION     AL1[illegible]442.1  GI:[illegible]
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; [illegible]
            Neoptera; Endopterygota; Diptera; Nematocera; [illegible]
            Anopheles.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible]-FEB-[illegible]) Genoscope - Centre National de Sequencage :
            BP [illegible] EVRY cedex - FRANCE [illegible]
            Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to [illegible])
  AUTHORS   [illegible]
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible]) [illegible] Institut Pasteur, [illegible]
            Roux, Paris 75[illegible], France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and [illegible] of Insects, Institut
            Pasteur.
```

```
FEATURES             Location/Qualifiers
     source          1..857
                     /organism "Anopheles gambiae"
                     /strain "PEST"
                     /db_xref "taxon:7165"
                     /clone "24H10"
                     /clone_lib "Notredame1"
                     /note "end : SP6"
BASE COUNT      270 a    151 c    142 g    290 t      4 others
ORIGIN

  Query Match            92.4%;  Score 19.4;  DB 17;  Length 857;
  Best Local Similarity  95.2%;  Pred. No. 2.7e+03;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGTTAAAAGT 21
              |||||||||||||||||| ||
Db        601 ATGAAAAGGAAAGTTAAAAAT 621


RESULT 2
CNS01NUR
LOCUS       CNS01NUR                 863 bp    DNA     linear   GSS 14-JUN-2000
DEFINITION  Anopheles gambiae GSS SP6 end of clone 24005 of Notredame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL152708
VERSION     AL152708.1  GI:7[illegible]
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 863)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1..863
                     /organism "Anopheles gambiae"
                     /strain "PEST"
                     /db_xref "taxon:7165"
                     /clone "24005"
                     /clone_lib "Notredame1"
                     /note "end : SP6"
BASE COUNT      275 a    153 c    142 g    291 t      2 others
ORIGIN

  Query Match            92.4%;  Score 19.4;  DB 17;  Length 863;
  Best Local Similarity  95.2%;  Pred. No. 2.7e+03;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGTTAAAAGT 21
              |||||||||||||||||| ||
Db        578 ATGAAAAGGAAAGTTAAAAAT 598


RESULT 3
AA514077
LOCUS       AA514077                 211 bp    mRNA    linear   EST 10-JUL-1997
DEFINITION  MCZAS032.ACS S. mansoni cercarial lambda Zap Schistosoma mansoni
            cDNA clone AS-32 5', mRNA sequence.
ACCESSION   AA514077
VERSION     AA514077.1  GI:225[illegible]
KEYWORDS    EST.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 211)
  AUTHORS   Karim,A.M.
  TITLE     Schistosome Gene Identification
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Karim,A.M.
            Dr.Amr Karim, Department of Biochemistry
            Faculty of Science, Ain Shams University
            Box 11[illegible], El Khalifa E. Maamoun St., Abbasia, Cairo, Egypt
            Tel: (202)2853561
            Fax: (202)2853561
            Email: Hady@asunet.shams.eun.eg
            Seq primer: SK
            High quality sequence stop: 211.
FEATURES             Location/Qualifiers
     source          1..211
                     /organism "Schistosoma mansoni"
                     /db_xref "taxon:6183"
                     /clone "AS-32"
                     /clone_lib "S. mansoni cercarial lambda Zap"
                     /sex "Mixed"
                     /dev_stage "cercarial"
                     /note "Library provided by Dr. M. Saber TBRI,Cairo,Egypt."
BASE COUNT       61 a     36 c     51 g     63 t
ORIGIN

  Query Match            87.6%;  Score 18.4;  DB [illegible];  Length 211;
  Best Local Similarity  95.0%;  Pred. No. [illegible];
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 TGAAAAGGAAAGTTAAAAGT 21
              |||||||||||| |||||||
Db        153 TGAAAAGGAAAGTTAAAAGT 172


RESULT 4
AV276062/c
LOCUS       AV276062                 225 bp    mRNA    linear   EST 1[illegible]-NOV-[illegible]
DEFINITION  AV276062 RIKEN full-length enriched, adult male testis (CH[illegible]) Mus
            musculus cDNA clone 4933432H2[illegible] 3', mRNA sequence.
ACCESSION   AV276062
VERSION     AV276062.1  GI:6264099
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Ebata,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kai,K., Kaiho,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Konno,H., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,R., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Toyoda,N., Tsuneda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,I.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 23[illegible]-04[illegible], Japan
            Tel: 81-45-503-9222
```

```
[illegible]
[illegible] Y., Hayashizaki
[illegible] sequencing using RNA
[illegible] (1998)
[illegible] Kawai,J.,
[illegible] Muramatsu,M., [illegible]
[illegible] plasmid preparation
[illegible] Methods Enzymol. [illegible]
[illegible]
[illegible] NA library was
[illegible] Encyclopedia
[illegible] Riken.
[illegible] Laboratory in
[illegible] Research in Riken.
[illegible] 1st strand cDNA was
[illegible] VN [illegible] cDNA was
[illegible] activated reverse
[illegible] for full length by
[illegible] Second strand cDNA was prepared with the
[illegible] cDNA
[illegible] sites. Vector: [illegible]
[illegible] excision from lambda
[illegible] end: BamHI."
[illegible]

[illegible] Gaps [illegible]

[illegible] SYSTEMS [illegible]

[illegible]

  AUTHORS   Shoemaker,R., [illegible]
            Wylie,T., Underwood,K., [illegible]
            [illegible] Harvey,N., [illegible]
            [illegible]
            .R., Waterston,R. and Wilson,R.K.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/USDA [illegible]
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box [illegible] USA
            Tel: 314 286 18[illegible]
            Fax: 314 286 18[illegible]
            Email: est@watson.wustl.edu
            This clone is available [illegible]
            South Memorial Parkway Huntsville, AL [illegible]
            [illegible]
            High quality sequence stop: [illegible]
FEATURES             Location/Qualifiers
     source          1..444
                     /organism="[illegible]"
                     /db_xref="taxon:[illegible]"
                     /note="[illegible] SYSTEMS [illegible]"
                     /[illegible]
                     /tissue_type="[illegible]"
                     /[illegible]
                     /note="Vector: [illegible]
                     Xho I. The cDNA library was [illegible] mRNA [illegible]
                     [illegible]
                     [illegible] cDNA was
                     synthesized from mRNA [illegible]
                     poly(dT) sequence with a Xho I [illegible]
                     adapters were [illegible] cDNA fragments
                     followed by Xho I [illegible] cDNA fragments were
                     directionally [illegible]
                     [illegible] cDNA fragments
                     were transformed [illegible]
                     library was [illegible]
                     Shoemaker."
BASE COUNT     [illegible] a     [illegible] c     [illegible] g     [illegible] t
ORIGIN

  Query Match             [illegible]%;  Score [illegible];  Length 444;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches   19;  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy          1 ATGAAAAGGAAATTAAAAG 2[illegible]
              [illegible]
Db        [illegible] ATGAAAAGAAAATTAAAA [illegible]


RESULT [illegible]
AL[illegible]
LOCUS       AL[illegible]  [illegible] bp  [illegible]NA  [illegible]
DEFINITION  [illegible]
            sapiens [illegible] NA [illegible]
ACCESSION   AL[illegible]
VERSION     AL[illegible]
KEYWORDS    [illegible]
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; [illegible]
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Mahairas,G.G., Wallace,[illegible]
            Keller,A., Shaker,[illegible]
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. [illegible]
  MEDLINE   [illegible]
COMMENT     Contact: Mahairas GG, Wallace [illegible]
```

```
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616 3618
            Fax: (206) 616 3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3080  row: P  column: 21
            Class: BAC ends
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1..457
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "Plate=3080 Col=21 Row=P"
                     /clone_lib "CIT Approved Human Genomic Sperm Library D"
                     /sex "male"
                     /note "Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E.Coli DH10B"
BASE COUNT      135 a     92 c     72 g    158 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 17;  Length 457;
  Best Local Similarity   95.0%;  Pred. No. 6.4e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGTAAAAG 20
              |||| ||||||||||||||
Db        163 ATGAACAGGAAAGTAAAAG 144


RESULT 7
BB665103/c
LOCUS       BB665103                 649 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB665103 RIKEN full-length enriched, 0 day neonate eyeball Mus
            musculus cDNA clone E130103M11 5', mRNA sequence.
ACCESSION   BB665103
VERSION     BB665103.1  GI:16498842
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kondo
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. : 10 (10), 1617-1630 (2000)
             Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. :
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. : 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1..649
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone "E130103M11"
                     /clone_lib "RIKEN full-length enriched, 0 day neonate
                     eyeball"
                     /tissue_type "eyeball"
                     /dev_stage "0 day neonate"
                     /lab_host "DH10B"
                     /note "Site_1: SalI; Site_2: BamHI. cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'-
                     GAGAGAGAGAGAGCGGCCGCAATTGAGTTTTTTTTTTTTTTTTVN-3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'-
                     GAGAGAGAGATCTCGAGTTAATTAAATTAATC...-3']. cDNA
                     was cleaved with BamHI and XhoI. Vector is a modified
                     pBluescript KS(+) after bulk excision from lambda ZAP II."
BASE COUNT      160 a    151 c    196 g    141 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 1;  Length 649;
  Best Local Similarity   95.0%;  Pred. No. 6.5e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGTAAAAG 20
              |||||||||||| ||||||
Db        638 ATGAAAAGGAAAGTAAAAAG 619


RESULT 8
AZ992563
LOCUS       AZ992563                 654 bp    DNA     linear   GSS 17-APR-2001
DEFINITION  2M0277G14F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0277G14 F, DNA sequence.
ACCESSION   AZ992563
VERSION     AZ992563.1  GI:13885790
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
```

```
[illegible]
"V [illegible] cDNA from M.
[illegible] from the Jackson
[illegible] The cDNA
[illegible] The sheared cDNA
[illegible] polymerase and T4
[illegible] nucleotides were
[illegible] express. The
[illegible] selected [illegible]
[illegible] derivative
[illegible] The vector was ligated
[illegible] insert adaptors and
[illegible] cDNA was annealed to
[illegible] (Stratagene) cells
[illegible]

[illegible] Length 664;
[illegible] Gaps [illegible]

[illegible] Vertebrata; Euteleostomi;
[illegible] Primates; [illegible] Homo.
[illegible]
[illegible] Health, Mammalian Gene Collection (MGC)
[illegible]

                     Clone distributed by MGC [illegible]
                     found through the I.M.A.G.E. [illegible]
                     http://[illegible]
                     Plate: LLAM[illegible]
                     High quality sequence [illegible]
FEATURES             Location/Qualifiers
     source          [illegible]
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:[illegible]"
                     /clone_lib="NIH_MGC_[illegible]"
                     /cell_line="MGC[illegible]"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. [illegible]
                     Average insert size [illegible] kb; Library prepared by Life
                     Technologies."
BASE COUNT     [illegible]
ORIGIN

  Query Match             [illegible]  Score [illegible]  DB [illegible]  Length [illegible]
  Best Local Similarity   [illegible]  Pred. No. [illegible]
  Matches [illegible]  Conservative [illegible]  Mismatches [illegible]  Indels [illegible]  Gaps [illegible]

Qy          [illegible] TAAAATAAATTAAAAT [illegible]
                 |                |
Db      [illegible] TAAATTAAATTAAAAT [illegible]


RESULT 1[illegible]
BE5[illegible]
LOCUS       BE5[illegible]  [illegible] bp    mRNA    [illegible] EST [illegible]
DEFINITION  [illegible] NIH_MGC_[illegible] Homo sapiens cDNA clone IMAGE:[illegible]
            mRNA sequence.
ACCESSION   BE5[illegible]
VERSION     BE5[illegible]
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distributed by MGC [illegible]
            found through the I.M.A.G.E. [illegible]
            http://[illegible]
            Plate: LLAM[illegible]
            High quality sequence [illegible]
FEATURES             Location/Qualifiers
     source          1..887
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:[illegible]"
                     /clone_lib="NIH_MGC_[illegible]"
                     /cell_line="MGC[illegible]"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: [illegible]
                     Average insert size 1.5 kb; Library prepared by Life
                     Technologies."
BASE COUNT     26[illegible] a    17[illegible] c    [illegible] g    [illegible] t
ORIGIN

  Query Match             8[illegible]%;  Score [illegible];  DB [illegible];  Length 887;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible]
```

```
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 TGAAAAGGAAAGCTAAAAGT 21
              |||| |||||||||||||||
Db        285 TGAAATGGAAAGCTAAAAGT 304


RESULT 11
BB326258/c
LOCUS       BB326258                 281 bp    mRNA    linear   EST 11-JUL-2000
DEFINITION  BB326258 RIKEN full-length enriched, 4 days neonate male adipose
            Mus musculus cDNA clone B430206I24 3', mRNA sequence.
ACCESSION   BB326258
VERSION     BB326258.1  GI:9035021
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kojima,M., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,K., Yasunishi,A., Yokota,I., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1..281
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="B430206I24"
                     /clone_lib="RIKEN full-length enriched, 4 days neonate
                     male adipose"
                     /sex="male"
                     /tissue_type="adipose"
                     /dev_stage="4 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'-
                     GAGAGAGAGAAGGATCAAGATTTTTTTTTTTTTTTTVN-3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 1.5 and subtraction to Rot = 22.4. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5'-GAGAGAGAGACTCGAGTTAATTAAATTAAT[illegible]
                     3'], cDNA was cleaved with XhoI and BamHI. Vector is a
                     modified pBluescript KS(-) after mass excision from lambda
                     FLC-1."
BASE COUNT       57 a     67 c     40 g    116 t      1 others
ORIGIN

  Query Match             85.7%;  Score 18;  DB 8;  Length 281;
  Best Local Similarity   100.0%;  Pred. No. [illegible];
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GAAAAGGAAAGCTAAAAG 20
              ||||||||||||||||||
Db        170 GAAAAGGAAAGCTAAAAG 153


RESULT 12
BB552761/c
LOCUS       BB552761                 306 bp    mRNA    linear   EST 05-APR-2001
DEFINITION  BB552761 RIKEN full-length enriched, 2 days pregnant adult female
            ovary Mus musculus cDNA clone E330021H21 3', mRNA sequence.
ACCESSION   BB552761
VERSION     BB552761.1  GI:9699127
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kojima,M., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,K., Yasunishi,A., Yokota,I., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
```

```
[illegible]
               /organism "Mus mus[illegible]"
               /sex "Female"
               /lab_host "E. [illegible] strain XL[illegible] resistant[illegible]"
               /note "Vector: pWI[illegible] [illegible] DNA [illegible]
               [illegible] was [illegible]
               [illegible] Mouse [illegible]NA [illegible]
               (http://www.[illegible]
               was hydrodynamically sheared by repeated passage [illegible]
               [illegible] at constant [illegible]. The sheared DNA
               was blunt end repaired with T4 DNA polymerase and T4
               polynucleotide kinase. Adaptors [illegible] were
               [illegible]
               adaptored DNA was [illegible]
               [illegible] kb [illegible] preparative [illegible]
               electrophoresis. Vector DNA was prepared from a derivative
               of pWI[illegible] copy number
               [illegible] derivative [illegible]. The vector was [illegible]
               with adaptors [illegible] adaptors and
               purified. The sheared [illegible] DNA was [illegible]
               adaptored vector DNA [illegible]
               chemically competent E. [illegible] XL[illegible] (Stratagene, [illegible])
               and selected for ampicillin resistance."
BASE COUNT     [illegible]
ORIGIN

  Query Match            [illegible]
  Best Local Similarity  [illegible]
  Matches   18; Conservative [illegible] Mismatches [illegible] Gaps [illegible]

Qy         4 AAAA[illegible]AAA[illegible]AAAA[illegible]

Db        [illegible]4 AAAA[illegible]AAA[illegible]AAAA[illegible]


RESULT 14
BH[illegible]
LOCUS       BH[illegible]   DNA [illegible]
DEFINITION  [illegible]
            DNA sequence.
ACCESSION   BH[illegible]
VERSION     BH[illegible]
KEYWORDS    GSS.
SOURCE      [illegible]
  ORGANISM  [illegible]
            [illegible]
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   [illegible]
  TITLE     [illegible]
  JOURNAL   Genomics [illegible]
  MEDLINE   [illegible]
COMMENT     Other_GSSs: [illegible]
            Contact: [illegible]
            INSERM [illegible]
            Institut Pasteur [illegible]
            [illegible] Professeur A. [illegible]
            Tel: [illegible]
            Fax: [illegible]
            Email: [illegible]pasteur[illegible]
            [illegible] AA [illegible]
            [illegible]
            Sequencing: [illegible] reverse [illegible]
            Class: [illegible]
            High quality sequence stop: [illegible]
```

```
FEATURES             Location/Qualifiers
     source          1..793
                     /organism="Schistosoma mansoni"
                     /strain="Puerto Rican"
                     /db_xref="taxon:6183"
                     /clone="032B24"
                     /clone_lib="SmBAC1"
                     /sex="mixed"
                     /dev_stage="cercariae"
                     /lab_host="Biomphalaria glabrata"
                     /note="Vector: pBeloBAC 11; Site_1: Hind III; Partially
                     Hind III digested and size selected S. mansoni  cercarial
                     DNA was ligated into Hind III digested pBeloBAC 11 vector
                     and used to transform E. coli DH10B. The complete library
                     contains 23808 clones from 4 independent
                     sizing ligation transformations.  Average insert size
                     ranges from 70 127 kb and genome coverage is 7.9 fold."
BASE COUNT      237 a    173 c    126 g    241 t     16 others
ORIGIN

  Query Match             85.7%;  Score 18;  DB 17;  Length 793;
  Best Local Similarity   90.0%;  Pred. No. 8.2e+03;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAAAAGGAAAGCTAAAAG 20
              ||||||||||||||| ||||
Db        692 ATGAAAAGGAAAGCRAAAAG 673


RESULT 17
CNS07RBF/c
LOCUS       CNS07RBF                 793 bp    DNA     linear   GSS 10-OCT-2001
DEFINITION  T3 end of clone 032DA12 of library SmBAC1 from strain Puerto Rican
            of Schistosoma mansoni, genomic survey sequence.
ACCESSION   AL623891
VERSION     AL623891.1  GI:16038035
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 793)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
REFERENCE   2  (bases 1 to 793)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              Web : www.genoscope.cns.fr)
COMMENT     Partially Hind III digested and size selected S. mansoni cercarial
            DNA was ligated into Hind III digested pBeloBAC 11 vector and used
            to transform E. coli DH10B. The complete library contains 23808
            clones from 4 independent sizing ligation transformations. Average
            insert size ranges from 70 127 kb and genome coverage is 7.9 fold.
FEATURES             Location/Qualifiers
     source          1..793
                     /organism="Schistosoma mansoni"
                     /strain="Puerto Rican"
                     /db_xref="taxon:6183"
                     /clone="032DA12"
                     /clone_lib="SmBAC1"
                     /note="end : T3"
BASE COUNT      237 a    173 c    126 g    241 t     16 others
ORIGIN

  Query Match             85.7%;  Score 18;  DB 17;  Length 793;
  Best Local Similarity   90.0%;  Pred. No. 8.2e+03;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels

Qy          1 ATGAAAAGGAAAGCTAAAAG 20
              ||||||||||||||| ||||
Db        692 ATGAAAAGGAAAGCRAAAAG

Search completed: November 4, 2002, 22:14:38
Job time : 43.5704 secs
```

ALIGNMENTS

RESULT 1
AAW2...
ID AAW2... standard; Protein; ... AA.
XX
AC AAW2...;
XX
DT 16-FEB-1998 (first entry)
XX
DE Murine granulocyte peptide A precursor (mature form MGP A).
XX
KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;
KW fungicide; antiprotozoal; protozoacide; antiviral; virucide;
KW murine granulocyte peptide A; MGP A; preservative; sepsis;
KW endotoxaemia; mouse.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Peptide ...
FT /label= Mat_peptide
FT /note= "MGP A antimicrobial peptide ..."
XX
PN WO9724...-A1.
XX
PD 21-AUG-1997.
XX
PF 13-FEB-1997; ...
XX
PR 16-FEB-1996; ...
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Selsted ME;

```
XX
DR   WPI; 1997-424753/39.
DR   N-PSDB; AAT78510.
XX
PT   Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -
PT   useful therapeutically, as preservatives for food, in water
PT   treatment and in agriculture
XX
PS   Claim 9; Fig 5; 56pp; English.
XX
CC   This protein comprises the precursor of a novel, claimed
CC   antimicrobial peptide from murine neutrophils, designated murine
CC   granulocyte peptide A or MGP A (see AAW23725).  Its amino acid
CC   sequence was deduced from a cDNA clone (see AAT78510) obtained from
CC   murine bone marrow.  MGP A and the bovine homologue, BGP-A (see
CC   AAW23724), exhibit activity against Gram positive and Gram-negative
CC   bacteria, fungi and viruses, specifically Staphylococcus aureus,
CC   Escherichia coli, Candida albicans, Salmonella typhimurium and C.
CC   neoformans (claimed).  They can be used in human or veterinary
CC   medicine (particularly to treat disorders associated with
CC   lipopolysaccharides, e.g. sepsis and endotoxaemia) or as
CC   preservatives in food products or in water supplies (claimed).
CC   They can also be applied to crops to reduce post-harvest spoilage
CC   or expressed in transgenic plants to increase their disease
CC   resistance.  They have low immunogenicity. Carboxamidated analogues
CC   of MGP A and BGP A may also be used.
XX
SQ   Sequence   181 AA;

  Query Match             100.0%;  Score 86;  DB 18;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGFLRSNYLVKGHRDV 16
              ||||||||||||||||
Db        144 RGFLRSNYLVKGHRDV 159


RESULT 2
AAY00776
ID   AAY00776 standard; Protein; 182 AA.
XX
AC   AAY00776;
XX
DT   18 MAY 1999  (first entry)
XX
DE   Mouse taq7 clone protein sequence.
XX
KW   Taq7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
KW   melanoma; leukaemia; apoptosis inducer; mouse.
XX
OS   Mus sp.
XX
PN   WO9902686-A1.
XX
PD   21-JAN-1999.
XX
PF   15-JUL-1998;   98WO-EP04287.
XX
PR   11-JUL-1997;   97US-0893764.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM INT GMBH.
XX
PI   Georgiev G,  Kiselev S,  Ostermann E,  Proklarchouk E;
XX
DR   WPI; 1999-120887/10.
DR   N-PSDB; AAX21814.
XX
PT   New nucleic acid encoding taq7 - used to inhibit tumour growth and
PT   induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
PT   leukaemia
XX
PS   Claim 19; Fig 1; 138pp; English.
XX
CC   This sequence is the murine taq7 of the invention. Cells containing
CC   the taq7 DNA sequence are used to express recombinant taq7. Taq7 is used
CC   to produce and purify antibodies; to inhibit growth of mammalian tumours,
CC   especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
CC   prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
CC   and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
CC   osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular
CC   weight marker. The taq7 polypeptide inhibits tumour growth and induces
CC   apoptosis. The taq7 coding sequences are also useful as probes for gene
CC   mapping and detection of taq7 gene expression, and as primers. Antibodies
CC   against taq7 are used as reagents for detecting taq7, as an antagonist of
CC   taq7; for isolating taq7 and therapeutically to inhibit or delay tumour
CC   metastasis.
XX
SQ   Sequence   182 AA;

  Query Match             100.0%;  Score 86;  DB 20;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RGFLKSNYEVKGHRDV 16
              ||||||||||||||||
Db        145 RGFLKSNYEVKGHRDV 160


RESULT 3
AAW23722
ID   AAW23722 standard; Protein; 190 AA.
XX
AC   AAW23722;
XX
DT   18-FEB-1998  (first entry)
XX
DE   Bovine granulocyte peptide A precursor (antimicrobial BGP A).
XX
KW   Antimicrobial peptide; antibiotic; antibacterial; antifungal;
KW   fungicide; antiprotozoal; protozoacide; antiviral; virucide;
KW   bovine granulocyte peptide A; BGP A; preservative; sepsis;
KW   endotoxaemia; cattle.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= Sig_peptide
FT   Peptide         22..177
FT                   /label= Propeptide
FT   Peptide         178..190
FT                   /label= Mat_peptide
FT                   /note= "BGP A antimicrobial peptide (claim 2)"
XX
PN   WO9729765-A1.
XX
PD   21-AUG-1997.
XX
PF   13-FEB-1997;   97WO-US02219.
XX
PR   16-FEB-1996;   96US-0011834.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Selsted ME;
XX
DR   WPI; 1997-424753/39.
DR   N-PSDB; AAT78509.
XX
PT   Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -
PT   useful therapeutically, as preservatives for food, in water
PT   treatment and in agriculture
XX
PS   Claim 8; Fig 4; 56pp; English.
XX
```

[illegible]

```
CC   [illegible] metastasis.
XX
SQ   Sequence [illegible] AA;

  Query Match              [illegible];  Score [illegible];
  Best Local Similarity    [illegible];  Pred. No. [illegible];
  Matches [illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy          [illegible] RGHLRSNYFVKGHRIV [illegible]
Db      [illegible] GALRSNYVIKGHRIV [illegible]


RESULT [illegible]
AAB24[illegible]
ID   AAB24[illegible] standard; [illegible] AA.
XX
AC   AAB24[illegible];
XX
DT   [illegible] JAN [illegible] (first entry)
XX
DE   Human [illegible] protein sequence SEQ ID NO:[illegible].
XX
KW   Human; [illegible]
KW   [illegible]
XX
OS   Homo sapiens.
XX
PN   WO[illegible]-A1.
XX
PD   [illegible] SEP [illegible].
XX
PF   [illegible] DEC [illegible];  [illegible]WO-US[illegible].
XX
PR   [illegible] MAR [illegible];  [illegible]US[illegible].
PR   [illegible] SEP [illegible];  [illegible]US[illegible].
PR   [illegible] [illegible];  [illegible]US[illegible].
PR   [illegible] NOV [illegible];  [illegible]US[illegible].
PR   [illegible] DEC [illegible];  [illegible]US[illegible].
XX
PA   (GETH ) GENENTECH INC.
XX
PI   [illegible] A, [illegible] MA, [illegible]
XX
DR   WPI; [illegible].
DR   N-PSDB; AA[illegible].
XX
PT   Antibodies specific for [illegible] polypeptides [illegible]
PT   the growth of tumors in mammals, [illegible]
PT   polypeptide activity or expression [illegible]
XX
PS   Claim [illegible]; [illegible]pp; English.
XX
CC   The present invention describes [illegible]
CC   protein (I) selected from [illegible]
CC   [illegible]
CC   [illegible]
CC   anticancer activity and can be used [illegible]
CC   detecting complex formation when [illegible]
CC   cells. Increased expression [illegible]
CC   to diagnose tumors. Agents which [illegible] (I),
CC   especially the antibodies, [illegible] which
CC   hybridises [illegible] (I), can be used [illegible]
CC   preferably by [illegible] cell death. Methods [illegible]
CC   can be used to identify compounds which [illegible]
CC   of (I). AA[illegible] to AA[illegible] represent [illegible]
CC   probes used in examples from the present [illegible] human [illegible]
CC   sequences. AA[illegible] to AA[illegible] and AAB[illegible] to AAB[illegible] represent human
CC   [illegible] polynucleotide and protein sequences [illegible] the examples [illegible]
CC   the present invention.
```

```
XX
SQ   Sequence   196 AA;

  Query Match             21.4%;  Score 64;  DB 21;  Length 196;
  Best Local Similarity   75.0%;  Pred. No. 0.0049;
  Matches   12;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 EQFLRSNYEVKGHRCV 16
                : ||||:||:||||
Db        158 QQALRSNYVLKGHRCV 173


RESULT 6
AAB25583
ID   AAB25583 standard; Protein; 196 AA.
XX
AC   AAB25583;
XX
DT   27-NOV-2000  (first entry)
XX
DE   Bt447 protein encoded by human secreted protein gene #8.
XX
KW   Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
KW   antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
KW   anticancer; vulnerary; antiviral; antibacterial; antifungal;
KW   immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW   multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW   Crohn's disease; nephritis; hyperproliferative disorder;
KW   cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW   melanoma; lymphoma; wound healing; human.
XX
OS   Homo sapiens.
XX
PN   WO200029435-A1.
XX
PD   25-MAY-2000.
XX
PF   27-OCT-1999;   99WO-US25031.
XX
PR   28-OCT-1998;   98US-0105971.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ni J,  Ruben SM,  Olsen HS,  Young PE,  Kenny JJ,  Moore PA,  Wei Y,
PI   Greene JM;
XX
DR   WPI; 2000-387742/33.
DR   N-PSDB; AAA80613.
XX
PT   Isolated nucleic acid molecules encoding human secreted proteins are
PT   used for the prevention, amelioration and treatment of autoimmune,
PT   inflammatory, hyperproliferative and cardiovascular disorders, cancer,
PT   wounds, and infectious diseases
XX
PS   Claim 1; Figure 34; 803pp; English.
XX
CC   The present invention relates to 12 secreted human proteins and the
CC   nucleotide sequences encoding them. The polynucleotide sequences given
CC   in AAA80606-AA80623 encode the 12 secreted protein sequences given in
CC   AAB25576-B25597. The human secreted proteins have various activities
CC   dependent on the tissues in which they are expressed. Examples of the
CC   activities of the proteins include: immunosuppressant;
CC   anti-inflammatory; antiarthritic; antirheumatic, dermatological;
CC   antiproliferative; antiarteriosclerotic; anticancer; vulnerary;
CC   antiviral; antibacterial; and antifungal activity. The proteins,
CC   polypeptides, agonists and antagonists may be used to treat prevent
CC   [illegible] examples of
CC   which include: immune disorders e.g. Addison's disease, rheumatoid
CC   arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
CC   e.g. inflammatory bowel disease, Crohn's disease and nephritis;
CC   hyperproliferative disorders such as paraproteinaemias and purpura;
CC   cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
CC   cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
CC   sequences may also be used in wound healing and the treatment of
CC   infectious diseases. The human secreted protein gene #8 and protein
CC   sequences are represented in sequences AAA80613 and AAB25583. Sequences
CC   AAA80662-A80663 represent genes related to the secreted protein gene#8
XX
SQ   Sequence   196 AA;

  Query Match             21.4%;  Score 64;  DB 21;  Length 196;
  Best Local Similarity   75.0%;  Pred. No. 0.0049;
  Matches   12;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 EQFLRSNYEVKGHRCV 16
                : ||||:||:||||
Db        158 QQALRSNYVLKGHRCV 173


RESULT 7
AAY96964
ID   AAY96964 standard; Protein; 196 AA.
XX
AC   AAY96964;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Chondrosarcoma peptidoglycan recognition protein like protein.
XX
KW   Peptidoglycan recognition protein like protein; PGRP; EGF; [illegible]
KW   chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
KW   tumour necrosis factor; TNF; anti-infectious; megakaryocyte; platelet;
KW   inhibitor; protein co-ordinate data.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= Signal peptide
FT   Protein         22..196
FT                   /label= Mature protein
FT   Domain          34..117
FT                   /label= EGFR like domain
FT   Region          1..6
FT                   /label= Antigenic region
FT   Region          20..29
FT                   /label= Antigenic region
FT   Region          33..43
FT                   /label= Antigenic region
FT   Region          63..79
FT                   /label= Antigenic region
FT   Region          99..112
FT                   /label= Antigenic region
FT   Region          133..146
FT                   /label= Antigenic region
FT   Region          160..165
FT                   /label= Antigenic region
FT   Region          168..181
FT                   /label= Antigenic region
FT   Region          190..196
FT                   /label= Antigenic region
XX
PN   WO200039327-A1.
XX
PD   06-JUL-2000.
XX
PF   22-DEC-1999;   99WO-US30736.
XX
PR   23-DEC-1998;   98US-0113809.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Young PE,  Olsen HS;
XX
DR   WPI; 2000-452413/39.
DR   N-PSDB; AAA51719.
```

[illegible]

```
PR   29-OCT-1998;   98US-0106248.
PR   29-OCT-1998;   98US-0106384.
PR   29-OCT-1998;   98US-0108[illegible].
PR   30-OCT-1998;   98US-0106464.
PR   02-NOV-1998;   98US-0106[illegible].
PR   0[illegible]-NOV-1998;   98US-0106[illegible].
PR   03-NOV-1998;   98US-0106[illegible].
PR   03-NOV-1998;   98US-0106919.
PR   03-NOV-1998;   98US-0106932.
PR   03-NOV-1998;   98US-0106934.
PR   10-NOV-1998;   98US-0107783.
PR   17-NOV-1998;   98US-0108775.
PR   17-NOV-1998;   98US-010877[illegible].
PR   17-NOV-1998;   98US-0108787.
PR   17-NOV-1998;   98US-0108788.
PR   17-NOV-1998;   98US-0108851.
PR   17-NOV-1998;   98US-0108802.
PR   17-NOV-1998;   98US-0108806.
PR   17-NOV-1998;   98US-0108807.
PR   17-NOV-1998;   98US-0108[illegible]7.
PR   17-NOV-1998;   98US-0108[illegible].
PR   18-NOV-1998;   98US-0108[illegible].
PR   18-NOV-1998;   98US-010884[illegible].
PR   18-NOV-1998;   98US-0108[illegible].
PR   18-NOV-1998;   98US-0108851.
PR   18-NOV-1998;   98US-0108852.
PR   18-NOV-1998;   98US-0108858.
PR   18-NOV-1998;   98US-0108904.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR   WPI; 20[illegible]-387671/20.
DR   N-PSDB; AAA37082.
XX
PT   New mammalian DNA sequences encoding transmembrane, receptor or
PT   secreted PRO polypeptides, useful for screening of potential peptide or
PT   small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS   Claim 12; Fig 122; 774pp; English.
XX
CC   AAA3[illegible] to AAA37144 encode the new isolated human transmembrane,
CC   receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC   transmembrane and receptor PRO proteins can be used for screening of
CC   potential peptide or small molecule inhibitors of the relevant
CC   receptor/ligand interactions. The polypeptides and nucleotide sequences
CC   encoding them have various industrial applications, including uses as
CC   pharmaceutical and diagnostic agents. AAA37145 to AAA3[illegible] represent
CC   PCR primers and hybridisation probes used in the isolation of the PRO
CC   polypeptides from the present invention.
XX
SQ   Sequence   196 AA;

  Query Match              74.4%;  Score 64;  DB 21;  Length 196;
  Best Local Similarity    75.0%;  Pred. No. 0.0039;
  Matches   12;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 RGHLRSNYEVKGHREV 16
                :  ||| :||||||
Db        158 QGALRSNYVLKGHRDV 173


RESULT 9
AAB66149
ID   AAB66149 standard; protein; 196 AA.
XX
AC   AAB66149;
XX
DT   02-APR-2001  (first entry)
XX
DE   Protein of the invention #61.
XX
KW   Secreted; transmembrane; gene therapy.
XX
OS   Unidentified.
XX
PN   WO200078961-A1.
XX
PD   28-DEC-2000.
XX
PF   18-FEB-2000; 2000WO-US04342.
XX
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   16-DEC-1999;   99WO-US30095.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Botstein D, Desnoyers L, [illegible] N, [illegible]
PI   Gao W, Goddard A, Godowski PJ, Grimaldi [illegible], Gurney AL, [illegible] K,
PI   Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, [illegible]
PI   Watanabe CK, Williams PM, Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and [illegible]
PT   useful as hybridization probes, in chromosome and gene mapping and [illegible]
PT   therapy.
XX
PS   Claim 1; Fig 122; 787pp; English.
XX
CC   The present invention relates to secreted and transmembrane proteins.
CC   These proteins and the DNA encoding them may be used as hybridization
CC   probes, in chromosome and gene mapping and in the generation of
CC   anti sense RNA and DNA. They may also be used used to generate either
CC   transgenic animals or knockout animals which are in turn useful for
CC   development and screening of therapeutically useful reagents.
CC   The nucleic acids may also be used in gene therapy.
XX
SQ   Sequence   196 AA;

  Query Match              74.4%;  Score 64;  DB 2[illegible];  Length 19[illegible];
  Best Local Similarity    75.0%;  Pred. No. [illegible];
  Matches   12;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 RGHLRSNYEVKGHREV 16
                :  ||| :||||||
Db        158 QGALRSNYVLKGHRDV 173


RESULT 10
ABP08874
ID   ABP08874 standard; Protein; 81 AA.
XX
AC   ABP08874;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:[illegible].
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
```

[illegible] degenerative, [illegible]

[illegible] human proteins [illegible] (see Table 1 [illegible] human [illegible]RFX [illegible] are useful for [illegible] RFX-associated [illegible] for treating [illegible] RFX polynucleotide [illegible] therapy. [illegible] can be used in the [illegible] of liver, [illegible] haemorrhage, [illegible] related to [illegible] systemic [illegible] cholesterol ester [illegible] disorders, infectious [illegible] rheumatoid [illegible] RFX proteins are also [illegible] osteoporosis. [illegible] and for the [illegible] liver fibrosis. [illegible] resulting from [illegible]

[illegible] part of the printed [illegible] directly from WIPO [illegible]

[illegible] AA

```
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200071[illegible]42 A2.
XX
PD   [illegible]-SEP-2[illegible].
XX
PF   2[illegible]-MAR-20[illegible]; [illegible]W[illegible]US[illegible].
XX
PR   2[illegible]-MAR-20[illegible]; [illegible]
PR   [illegible]
XX
PA   (PEKE-) [illegible] NY
XX
PI   Venter JC, Adams M, [illegible] FWL, Myers EW
XX
DR   WPI; 20[illegible]/[illegible].
DR   N-PSDB; ABL[illegible].
XX
PT   New isolated nucleic acid detection reagent [illegible]
PT   genes from Drosophila and for elucidating [illegible] and cell-cell
PT   interactions
XX
PS   Disclosure; SEQ ID NO [illegible]
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting [illegible] or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL[illegible]) expressed DNA
CC   sequences (ABL[illegible]840-ABL[illegible]175) and the encoded proteins
CC   (ABB57737-ABB7207[illegible]).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence [illegible] AA;

  Query Match [illegible]
  Best Local Similarity [illegible]
  Matches [illegible] Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

Qy          1 RGFLRSNYEVK[illegible] 14
                [illegible]
Db        [illegible]2 RGYLKENY[illegible] [illegible]


RESULT 12
AAR[illegible]282
ID   AAR[illegible] standard; [illegible] AA.
XX
AC   AAR[illegible];
XX
DT   31-MAY-1994  (first entry)
XX
DE   Helminth aminopeptidase H[illegible] [illegible]
XX
KW   Helminth; aminopeptidase; [illegible]
KW   alpha-amino acyl peptide hydrolase [illegible]
KW   antigen; vaccine; parasite; human; [illegible]
XX
OS   Haemonchus contortus.
XX
PN   WO9[illegible]2[illegible]4[illegible]-A.
XX
PD   25-NOV-1993.
XX
PF   07-MAY-1[illegible]; [illegible]W[illegible]GB[illegible]4[illegible].
XX
PR   08-MAY-1[illegible]; [illegible]2GB-[illegible].
XX
```

```
PA   (AGRI-) AGRIC & FOOD RES COUNCIL.
XX
PI   Graham M,  Knox DP,  Munn EA,  Newton SE,  Oliver JJ,
PI   Smith TS;
XX
DR   WPI; 1998-386574/48.
DR   N-PSDB; AAQ52491.
XX
PT   New DNA encoding amino peptidase from Haemonchus contortus and
PT   derived peptide(s) - useful in protective vaccines, for new
PT   vectors, transformed cells and oligosaccharide, for incorporation
PT   in virus or microbe
XX
PS   Claim 6; Page 78-80; 137pp; English.
XX
CC   The sequences given in AAR51280-82 represent variants of the helminth
CC   aminopeptidase H110D encoded by the genes H11-3, -2 and -1 respectively.
CC   H110D is a protein doublet which shows homolgy to a family of integral
CC   membrane aminopeptidases.  The differences between the coding sequences
CC   and the translated products, can be attributed to different mRNAs of
CC   the multigene family, and also to different variants of the H110D
CC   encoding sequence being present at different stages of the life
CC   cycle, or in strains differing in geographical origin.  Antigenic
CC   fragments of these aminopeptidases may be used in vaccines to stimulate
CC   immune response against helminth parasites in humans or other animals.
CC   The DNA sequences encoding these proteins may be incorporated into a
CC   virus or microbe and used in a similar manner.
XX
SQ   Sequence   977 AA;

  Query Match             54.7%;  Score 47;  DB 14;  Length 977;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 RGFLRSNYEVKCHRDV 16
                | : ||||:  | ::
Db        630 RAFFRSNYDANGWRNI 645


RESULT 13
ABG26884
ID   ABG26884 standard; Protein; 334 AA.
XX
AC   ABG26884;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #26875.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS91071.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 20; SEQ ID NO 41230; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states [illegible]
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or [illegible] activity
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG0[illegible] to ABG30[illegible] represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   334 AA;

  Query Match             53.5%;  Score 46;  DB 2[illegible];  Length 334;
  Best Local Similarity   58.3%;  Pred. No. [illegible];
  Matches    7;  Conservative    3;  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy          4 LRSNYEVKCHRD 15
              |   ::: ||||
Db        292 LRDNFDIVGHRD 303


RESULT 14
AAB73613
ID   AAB73613 standard; Protein; 432 AA.
XX
AC   AAB73613;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Bacillus sp. KSM KP43 origin of replication associated protein.
XX
KW   Origin of replication associated protein; [illegible] plasmid construction;
KW   expression vector; recombinant expression.
XX
OS   Bacillus sp. KSM KP43.
XX
PN   JP2001136986-A.
XX
PD   21-MAR-2001.
XX
PF   03-SEP-1999;   99JP-0250690.
XX
PR   03-SEP-1999;   99JP-0250690.
XX
PA   (KAOS ) KAO CORP.
XX
DR   WPI; 2001-297392/31.
DR   N-PSDB; AAH23948.
XX
PT   A plasmid vector used for the preparation of a useful protein or a
PT   useful peptide
XX
PS   Claim 1; Page 6-7; 9pp; Japanese.
XX
CC   The invention relates to a plasmid vector constructed from a DNA
CC   fragment from Bacillus sp. KSM KP43 (AAH[illegible]) which contains the
CC   origin of replication (ori). This vector can be used for the
```

[illegible]

... detection;

... polypeptides, ... colorectal cancers

... human ... (P), where ... the ... used as ... prevention ... disorders ... or deletions ... expressing ... associated ... culturing the cell ... prevention, diagnosis ... AAB ... to AAB37204 ...

... were

SEQ ID N ...
XX
SQ Sequence [illegible] AA;

Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible] Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

Qy 1 KGFLESNYFVCGHEL [illegible]
Db [illegible] RNYIKHFYI [illegible]

Search completed: November [illegible]
Job time : [illegible] secs

[illegible]

Description

Sequence 7x, Appl
Sequence 7x, Appl
Sequence 64, Appl
Sequence 64, Appl
Sequence 2x, Appl
Sequence 4x, Ap
Sequence 5, Appli
Sequence 4949, Ap
Sequence 1xxxx, A
Sequence 1x, Appl
Sequence 1xxxx, A
Sequence 2x, Appl
Sequence 2, Appl
Sequence 4, Appli
Sequence 5882, Ap
Sequence 13079, A
Sequence 58, Appl
Sequence 58, Appl
Sequence 4, Appli

[illegible]

ALIGNMENTS

RESULT 1
US-[illegible]
; Sequence 7x, Application US/[illegible]
; Patent No. US2[illegible]A1
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven [illegible]
;   APPLICANT: [illegible], Michael [illegible]
;   APPLICANT: [illegible], Raymond
;   TITLE OF INVENTION: [illegible] AND METHODS FOR THE [illegible] AND
;   TITLE OF INVENTION: THERAPY OF [illegible] INFECTION
;   NUMBER OF SEQUENCES: 7x
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SEED and BERRY LLP
;   STREET: 6300 Columbia Center, 701 Fifth Avenue
;   CITY: Seattle
;   STATE: Washington
;   COUNTRY: USA
;   ZIP: 98104
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.x
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/[illegible]
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: [illegible]
;   FILING DATE: [illegible]
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Maki, [illegible]
;   REGISTRATION NUMBER: [illegible]
;   REFERENCE/DOCKET NUMBER: [illegible]
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: [illegible]
;   TELEFAX: [illegible]
; INFORMATION FOR SEQ ID NO: 7x:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: [illegible] amino acids
;   TYPE: amino acid
;   STRANDEDNESS:

```
;       TOPOLOGY:   linear
;     MOLECULE TYPE:  protein
US-09-159-469-73

  Query Match              5.7%;  Score 97.5;  DB 10;  Length 590;
  Best Local Similarity   22.8%;  Pred. No. 0.14;
  Matches   68;  Conservative   54;  Mismatches  101;  Indels   75;  Gaps   16;

Qy      74 DMVV [illegible] NNWGVLL [illegible] LAYVKN [illegible] VVALAVVK [illegible]
Db     110 DMVVEKFAHL [illegible] VFP [illegible] AVLA [illegible]
Qy     112 [illegible] AM [illegible] ES [illegible]
Db     164 SYVNK [illegible] YLALAK [illegible]
Qy     167 KE [illegible] KVSVY [illegible] MNGMEYA [illegible] KEKG [illegible] WAL [illegible] W 211
Db     214 V [illegible] AV [illegible] KE [illegible] YRSA [illegible] AVIV 269
Qy     212 SNPNV [illegible] EMS [illegible] KSE [illegible]
Db     270 [illegible] ENEAL [illegible] KSE [illegible]
Qy     265 [illegible] AHL [illegible]
Db     330 [illegible] WH [illegible] KHL [illegible]

RESULT 2
US-09-798-042-73
; Sequence 73, Application US/09798042
; Patent No. US20020068343A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Houghton, Raymond L.
;  APPLICANT: McNeill, Patricia D.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;  TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
;  FILE REFERENCE: 210121.43902
;  CURRENT APPLICATION NUMBER: US/09/798,042
;  CURRENT FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 158
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 73
;   LENGTH: 590
;   TYPE: PRT
;   ORGANISM: Ehrlichia sp.
US-09-798-042-73

  Query Match              5.7%;  Score 97.5;  DB 10;  Length 590;
  Best Local Similarity   22.8%;  Pred. No. 0.14;
  Matches   68;  Conservative   54;  Mismatches  101;  Indels   75;  Gaps   16;

Qy      74 DMVV [illegible] NNWGVLL [illegible] LAYVKN [illegible] VVALAVVK [illegible]
Db     110 DMVVEKFAHL [illegible] VFP [illegible]
Qy     112 [illegible] AMINP [illegible] GE [illegible]
Db     164 SYVNK [illegible] YLALAK [illegible] VA [illegible]
Qy     167 KE [illegible] MNGMEYA [illegible] KEKG [illegible] WAL [illegible] W 211
Db     214 V [illegible] AV [illegible] YRSA [illegible]
Qy     212 SNPNV [illegible] EMS [illegible] KSE [illegible]
Db     270 [illegible] ENEAL [illegible] KSE [illegible]
Qy     265 [illegible] AHL [illegible]
Db     330 NNIPKEVIPVIVENVVEALNE [illegible] WHIMKVI [illegible] 387

RESULT 3
US-09-159-469-64
; Sequence 64, Application US/09159469
; Patent No. US20020064975A1
;  GENERAL INFORMATION:
;    APPLICANT: Reed, Steven G.
;    APPLICANT: Lodes, Michael J.
;    APPLICANT: Houghton, Raymond
;    TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
;    TITLE OF INVENTION: TREATMENT OF EHRLICHIA INFECTION
;    NUMBER OF SEQUENCES: 73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: SEED and BERRY LLP
;      STREET: 6300 Columbia Center, 701 Fifth Avenue
;      CITY: Seattle
;      STATE: Washington
;      COUNTRY: USA
;      ZIP: 98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/159,469
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 09/106,582
;      FILING DATE: 29-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Maki, David J.
;      REGISTRATION NUMBER: 31,392
;      REFERENCE/DOCKET NUMBER: 210121.43902
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 206-622-4900
;      TELEFAX: 206-682-6031
;  INFORMATION FOR SEQ ID NO: 64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 590 amino acids
;      TYPE: amino acid
;      STRANDEDNESS:
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-09-159-469-64

  Query Match              5.7%;  Score 97.5;  DB 10;  Length 590;
  Best Local Similarity   22.8%;  Pred. No. 0.16;
  Matches   68;  Conservative   54;  Mismatches  101;  Indels   75;  Gaps   16;

Qy      74 DMVV [illegible] NNWGVLL [illegible] LAYVKN [illegible] VVALAVVKS [illegible]
Db     110 DMVVEKFAHL [illegible]
Qy     112 [illegible] AM [illegible]
Db     164 SYVNK [illegible]
Qy     167 [illegible] WAL [illegible] W 211
Db     214 V [illegible] AVIV 269
Qy     212 [illegible]
Db     270 [illegible] ENEAL [illegible]
Qy     265 [illegible] 316
Db     330 NNIPKEVIPVIVENVVEALNE [illegible] WHIMKVI [illegible] 387
```

```
Db     [illegible] IFLKRM[illegible]EK [illegible]MAR[illegible]EW[illegible]MF[illegible]INYM[illegible]

Qy     [illegible] AS[illegible]IVKEFL MVV[illegible]NN W[illegible]SAR[illegible]AYVKN[illegible]VVA[illegible]AVVK[illegible]

Db     [illegible]KVV[illegible]SSNVW V[illegible]A[illegible]VY[illegible]YKHN[illegible]

Qy     [illegible] KYA[illegible] VKV[illegible]

Db     145 [illegible]LRYS[illegible]VS[illegible]IV[illegible]IVI[illegible]M[illegible]VIEMEA [illegible]EM[illegible]AF[illegible]VV[illegible]M[illegible]

Qy     [illegible] [illegible]NIKIMKEIKVSVYS[illegible]

Db     [illegible] A[illegible]KVI[illegible]N[illegible]VIKE[illegible]V[illegible]YYN[illegible]

RESULT [illegible]
US-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. US[illegible]A1
; GENERAL INFORMATION:
;  APPLICANT: [illegible] et al.
;  TITLE OF INVENTION: N[illegible] Ac[illegible]
;  FILE REFERENCE: [illegible]
;  CURRENT APPLICATION NUMBER: US/[illegible]
;  CURRENT FILING DATE: [illegible]
;  PRIOR APPLICATION NUMBER: [illegible]
;  PRIOR FILING DATE: [illegible]
;  PRIOR APPLICATION NUMBER: [illegible]
;  PRIOR FILING DATE: [illegible]
;  NUMBER OF SEQ ID NOS: [illegible]
;  SOFTWARE: PatentIn Ver. 2.[illegible]
; SEQ ID NO [illegible]
;   LENGTH: [illegible]
;   TYPE: PRT
;   ORGANISM: H[illegible] sapiens
US-[illegible]

  Query Match [illegible]
  Best Local Similarity [illegible]; Pred. No. [illegible]
  Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy     4[illegible] FLARD[illegible]INYV[illegible]VKF[illegible]MVV[illegible]NNW[illegible]AYVKN[illegible]

Db     18 FEAFE[illegible]FASG [illegible]YKYS[illegible] [illegible]IVL 4[illegible]

Qy     [illegible] VAFAVVK[illegible]

Db     48 ASKAMEES[illegible]SLE[illegible]MS[illegible]VYISK[illegible]AVVEY[illegible]EKEN[illegible]VN [illegible]

Qy     162 NIKIMKEIKV[illegible]VY[illegible] [illegible]WAL[illegible]IWENEN[illegible]

Db     [illegible] NIYVL[illegible]AKE[illegible]EK[illegible]KE[illegible]MM[illegible] [illegible]VIWV[illegible]AN[illegible]

Qy     [illegible] YIENISS[illegible] [illegible]NY[illegible]A[illegible]EMEIKA [illegible]KVKN[illegible]

Db     [illegible]KMS[illegible]N[illegible]SAKA[illegible]AE[illegible]KA[illegible]ME[illegible]

Qy     2[illegible] FYVKDLKV[illegible]Y [illegible]DKLSVS[illegible]SE[illegible]SESVEKVY [illegible]KAE[illegible]KEV [illegible]

Db     21[illegible] [illegible]YE[illegible]IAE[illegible] [illegible]KVHE[illegible]EKVEAE[illegible]EKKAL[illegible]

Qy     [illegible] KIEEKISIAE SE[illegible]NEVEKIE[illegible]EKE [illegible]

Db     2[illegible] KINKEIV[illegible]V[illegible] NEVLKA[illegible]KEE [illegible]

RESULT [illegible]
US-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. US[illegible]A1
; GENERAL INFORMATION:
;  APPLICANT: Br[illegible]
;  APPLICANT: Ka[illegible]
```

```
;  APPLICANT:  Yihong Xie
;  APPLICANT:  Janet Elizabeth Paulsen
;  TITLE OF INVENTION: NOVEL GLUTAMATE RECEPTOR MODULATORY PROTEINS AND
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: GNN 024
;  CURRENT APPLICATION NUMBER: US/10/027,924
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/257,589
;  PRIOR FILING DATE: 2000-12-22
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1263
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-10-027-924-5

  Query Match             [illegible];  Score [illegible];  DB 12;  Length 1263;
  Best Local Similarity   [illegible];  Pred. No. 1.7;
  Matches [illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy          8 ILFFILSIVLFAQEIDGLAEGSKPA-  -EPGELVLP PAELARL[illegible]NYVPYV 62
Db          2 VIIIIISVILLKEEVPGSAQSSEPPVVAHMPG[illegible]VHHQPI -  -  VEKV 52
Qy         63 [illegible]                                                        117
Db         53 [illegible]                                                        112
Qy        118 [illegible]                                                        177
Db        113 [illegible]                --CV[illegible]                          147
Qy        178 [illegible]                                                        240
Db        148 [illegible]                                                        173
Qy        241 [illegible]                                                        299
Db        174      SDF[illegible]                     V[illegible]               224
Qy        300 [illegible]                                                        [illegible]
Db        225 ESGMKAFFMSAKE[illegible]HL[illegible]                              264

RESULT 8
US-09-815-242-4949
; Sequence 4949, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
```

```
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4949
;   LENGTH: 764
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-815-242-4949

  Query Match             [illegible];  Score [illegible];  DB [illegible];  Length 764;
  Best Local Similarity   21.9%;  Pred. No. 1.4;
  Matches [illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy          8 ILFF[illegible]LLSIVLLAQ[illegible]EIDGLAEGSKPA-              [illegible]
Db         17 [illegible]                                                        76
Qy         28 [illegible]                                                        [illegible]
Db         77 [illegible]                                                        [illegible]
Qy         82 [illegible]                                                        [illegible]
Db        132 [illegible]                                                        [illegible]
Qy        141 [illegible]                                                        [illegible]
Db        185 [illegible]                                                        244
Qy        176 [illegible]                                                        [illegible]
Db        245 [illegible]                                                        [illegible]
Qy        221 [illegible]                                                        [illegible]
Db        288 [illegible]                                                        [illegible]
Qy        2[illegible] [illegible]  298
Db        3[illegible] [illegible]  349

RESULT 9
US-09-815-242-10593
; Sequence 10593, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
```

[illegible] SYSTEMS FOR MAKING AND [illegible]

; OTHER INFORMATION: Strain [illegible]
US-09-7[illegible]-24 [illegible]

Query Match [illegible]
Best Local Similarity [illegible]
Matches 86; Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

```
Qy      1 MKKKAKS[illegible]
Db     40 IKEEKKSV[illegible]
Qy     54 [illegible]
Db    100 [illegible]NFNL[illegible]
Qy     9[illegible] [illegible]
Db    157 PS---HHFKENHHN[illegible]AVVN[illegible]
Qy    147 [illegible]
Db    [illegible]
Qy    1[illegible] [illegible]KFK[illegible]
Db    [illegible]
Qy    24[illegible] ---KALLVSK[illegible]
Db    [illegible]
Qy    268 YDKLSVS[illegible]
Db    [illegible]
```

RESULT 1[illegible]
US-09-81[illegible]-242-1374[illegible]
; Sequence 1374[illegible], Application US/09815242
; Patent No. US2[illegible]A
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: [illegible], Kari [illegible]
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.[illegible]A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: 2000-05-[illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; NUMBER OF SEQ ID NOS: [illegible]
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1374[illegible]
; LENGTH: 4[illegible]
; TYPE: PRT

```
; ORGANISM: Salmonella typhi
US-09-815-242-13745

  Query Match            5.1%;  Score 88.5;  DB 10;  Length 433;
  Best Local Similarity  19.0%;  Pred. No. 0.76;
  Matches   59;  Conservative   58;  Mismatches  101;  Indels  103;  Gaps   12;

Qy     50 SIRLLLINT                            VLTVYLSAASIVKPEDMVVLSGINNWSVL  87
Db      2 NIKLDAGNI[illegible]NTALLL[illegible]INVVVLE  60

Qy     54                                    LILSAELQAYVENSVVALAVVKSLEKY  A 116
Db     61 AEE[illegible]ISNL[illegible]  12

Qy    117 GLIILGVRVLI[illegible]  157
Db    121 AGIV[illegible]IGK[illegible]  178

Qy    158 [illegible]AY  [illegible]  2[illegible]
Db    179 HM[illegible]SAQAANS[illegible]SVKA  231

Qy    206 WA[illegible]R[illegible]  257
Db    242 AK[illegible]VERI[illegible]VINY  29[illegible]

Qy    258 IFYVKDLEVLY 268
Db    291 YRLIKENRLLF 301

RESULT 12
US-09-887-586A-20
; Sequence 20, Application US/09887586A
; Patent No. US20020094556A1
; GENERAL INFORMATION:
;  APPLICANT: Chappell, Joseph
;  APPLICANT:  No. US20020094556A1I, Joseph P.
;  APPLICANT:  Starks, Courtney M
;  APPLICANT:  Mannar, Kathleen E.
;  TITLE OF INVENTION: SYNTHASES
;  FILE REFERENCE: 07678-025001
;  CURRENT APPLICATION NUMBER: US/09/887,586A
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: 09/398,395
;  PRIOR FILING DATE: 1999-09-17
;  PRIOR APPLICATION NUMBER: 60/130,628
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: 60/150,262
;  PRIOR FILING DATE: 1999-08-2[illegible]
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 20
;   LENGTH: 628
;   TYPE: PRT
;   ORGANISM: Abies grandis
US-09-887-586A-20

  Query Match            5.1%;  Score 88.5;  DB 10;  Length 628;
  Best Local Similarity  24.0%;  Pred. No. 1.2;
  Matches   79;  Conservative   44;  Mismatches   85;  Indels   99;  Gaps   18;

Qy     29 SKEAEPGELVLDFAELARD     PSSTRL       DLINYVDYVYS  GASSI-VKPE 73
Db    108 SM[illegible]RALGYVT[illegible]  167

Qy     74 DMVVL[illegible]NNWSVL[illegible]AYVENSVVALAVVKSLEKYA       [illegible]  126
Db    168 SVVIDL  [illegible]  22[illegible]

Qy    127 [illegible]MEEI  [illegible]  1[illegible]

Db    224   KASLIAFF             [illegible]  2[illegible]

Qy    185 [illegible]  24[illegible]
Db    266       DVLEY           GW        HTYIEELEAF         NY            [illegible]

Qy    245 [illegible]VVKEISG [illegible]
Db    286 [illegible]VWKESG 3[illegible]

RESULT 13
US-09-903-612-20
; Sequence 20, Application US/09903612
; Patent No. US20020094557A1
; GENERAL INFORMATION:
;  APPLICANT: Chappell, Joseph
;  APPLICANT:  No. US20020094557A1I, Joseph P.
;  APPLICANT:  Starks, Courtney M.
;  APPLICANT:  Mannar, Kathleen E.
;  TITLE OF INVENTION: SYNTHASES
;  FILE REFERENCE: 07678-025001
;  CURRENT APPLICATION NUMBER: US/09/903,612
;  CURRENT FILING DATE:  2001-07-11
;  PRIOR APPLICATION NUMBER: 09/398,395
;  PRIOR FILING DATE: 1999-09-17
;  PRIOR APPLICATION NUMBER: 60/100,993
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: 60/130,628
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: 60/150,262
;  PRIOR FILING DATE: 1999-08-2[illegible]
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 20
;   LENGTH: 628
;   TYPE: PRT
;   ORGANISM: Abies grandis
US-09-903-612-20

  Query Match            5.1%;  Score 88.5;  DB 1[illegible];  Length 628;
  Best Local Similarity  24.0%;  Pred. No. 1.2;
  Matches   79;  Conservative   44;  Mismatches   85;  Indels   99;  Gaps   18;

Qy     29 SKEAEPGELVLDFAELARD     PSSTRL          DLINYVDYVYS  GASSI-VKPE 73
Db    108 [illegible]  167

Qy     74 DMVVL[illegible]AYVENSVVALAVVKSLEKYA       [illegible]  126
Db    168 SVVIDL    [illegible]  22[illegible]

Qy    127 [illegible]  184
Db    224     KASLIAFF          [illegible]  2[illegible]

Qy    185 [illegible]  244
Db    266     DVLEY             GW        HTYIEELEAF         NY            [illegible]

Qy    245 [illegible]VVKEISG 292
Db    286 [illegible]VWKESG 3[illegible]4

RESULT 14
US-10-027-[illegible]-4
; Sequence 4, Application US/10027[illegible]
; Patent No. US20020142330A1
; GENERAL INFORMATION:
;  APPLICANT: Briann Gaither Bates
;  APPLICANT:  [illegible]
;  APPLICANT:  Yuhong Xie
```

; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; PRIOR APPLICATION NUMBER: [illegible]
; PRIOR FILING DATE: [illegible]
; NUMBER OF SEQ ID NOS: [illegible]
; SOFTWARE: FastSEQ for Windows Version [illegible]
; SEQ ID NO [illegible]
; LENGTH: [illegible]
; TYPE: PRT
; ORGANISM: [illegible]

Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible] Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

Qy 74 [illegible]
Db 491 [illegible]

Qy [illegible]
Db 547 [illegible]

Qy [illegible]
Db 597 [illegible]

Qy [illegible]
Db [illegible]

Qy [illegible]
Db [illegible]

Search completed: November [illegible]
Job time : [illegible] secs

ALIGNMENTS

RESULT [illegible]
US-[illegible]-4
; Sequence 4, Application US/[illegible]
; Patent No. [illegible]
; GENERAL INFORMATION:
;   APPLICANT: Blanco, David E.
;   APPLICANT: Miller, [illegible] N.
;   APPLICANT: Lovett, Michael A.
;   APPLICANT: Champion, Cheryl [illegible]
;   APPLICANT: [illegible]
;   TITLE OF INVENTION: [illegible] AND AMINO ACID SEQUENCES [illegible] A
;   TITLE OF INVENTION: [illegible] FAMILY OUTER MEMBRANE PROTEIN
;   NUMBER OF SEQUENCES: [illegible]
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson [illegible]
;   STREET: 4225 Executive Square, Suite [illegible]
;   CITY: La Jolla
;   STATE: California
;   COUNTRY: USA
;   ZIP: [illegible]
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.[illegible]
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: [illegible]
;   FILING DATE: [illegible]-JAN-[illegible]
;   CLASSIFICATION: [illegible]
;   ATTORNEY/AGENT INFORMATION:
;   NAME: [illegible] M.
;   REGISTRATION NUMBER: [illegible]
;   REFERENCE/DOCKET NUMBER: [illegible]
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: [illegible]
;   TELEFAX: [illegible]
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 24[illegible] amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-[illegible]-4

Query Match [illegible]
Best Local Similarity 2[illegible]%; Pred. No. [illegible]
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy [illegible] TSSIRG[illegible]INYV[illegible]YV YSGAS[illegible]VKF[illegible]MVV[illegible]NNW [illegible]V[illegible]AF[illegible]AYVEN[illegible]

Db [illegible]

Qy [illegible]

Db [illegible]

Qy [illegible]

Db [illegible] IDVVVA [illegible]

Qy [illegible]

Db 174 HIPQIS [illegible] RYLKLAQHLSPVGPRIRLSGSEVDFY [illegible] VFIDQFKALA 217

Qy 27[illegible] DSDID 280

Db 218 NMHID 222

RESULT 2
US [illegible] 2
; Sequence 2, Application US/[illegible]
; Patent No. 6013266
; GENERAL INFORMATION:
; APPLICANT: Segers, Ruud P.A.M.
; APPLICANT: Frey, Joachim
; TITLE OF INVENTION: Live attenuated Actinobacillus
; TITLE OF INVENTION: pleuropneumoniae
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo Nobel 6013266el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: [illegible]
; FILING DATE: 0[illegible] APR 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948 7400
; TELEFAX: (301) 948 9751
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1657 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US 09 057 570 2

Query Match 6.4%; Score 11[illegible]; DB [illegible]; Length 1657;
Best Local Similarity 21.5%; Pred. No. 0.048;
Matches 68; Conservative [illegible]; Mismatches 114; Indels 94; Gaps 17;

Qy [illegible]

Db [illegible]

Qy [illegible]

Db 914 [illegible] NW [illegible] 91[illegible]

Qy 132 [illegible]

Db [illegible]

Qy [illegible]

Db [illegible]

Qy [illegible]

Db [illegible]

Qy [illegible] KAVEIEKEIIAEKSE 322

Db 1[illegible] [illegible] 11[illegible]

RESULT 3
US 09 057 570 7
; Sequence 7, Application US/[illegible]
; Patent No. 6013266
; GENERAL INFORMATION:
; APPLICANT: Segers, Ruud P.A.M.
; APPLICANT: Frey, Joachim
; TITLE OF INVENTION: Live attenuated Actinobacillus
; TITLE OF INVENTION: pleuropneumoniae
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo Nobel 6013266el Patent Department
; STREET: 1300 Piccard Drive, Suite 2[illegible]
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #[illegible] (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/057,570
; FILING DATE: 09 APR 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948 7400
; TELEFAX: (301) 948 9751
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US 09 057 570 7

Query Match 6.4%; Score 110; DB 3; Length 1805;
Best Local Similarity 21.5%; Pred. No. [illegible];
Matches 68; Conservative 51; Mismatches 114; Indels 94; Gaps [illegible];

Qy [illegible]

Db 1009 [illegible] VAKAQGEYE [illegible]

Qy [illegible]

Db [illegible] NW [illegible]

Qy [illegible]

Db [illegible]

Qy [illegible]

```
Db    27[illegible] S[illegible]EV[illegible]VE[illegible]KN[illegible] [illegible] AYKAY[illegible] [illegible] 3[illegible]

Qy    26[illegible] EV[illegible]Y[illegible]K[illegible]SVS[illegible]S[illegible]SESVEKVYE [illegible]KAHE[illegible] 3[illegible]

Db    33[illegible] NN[illegible]SK[illegible]V[illegible] [illegible]KHE[illegible]KEK[illegible] 3[illegible]

RESULT [illegible]
US-09-29[illegible]-628-7[illegible]
; Sequence 7[illegible], Application US/[illegible]
; Patent No. 62[illegible]
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven [illegible]
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: [illegible], Raymond L.
;  APPLICANT: McNeill, Patricia D.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE [illegible]
;  TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
;  FILE REFERENCE: 21[illegible].4[illegible]
;  CURRENT APPLICATION NUMBER: US/[illegible]
;  CURRENT FILING DATE: 1999-04-2[illegible]
;  NUMBER OF SEQ ID NOS: [illegible]
;  SOFTWARE: FastSEQ for Windows Version [illegible]
; SEQ ID NO 7[illegible]
;   LENGTH: [illegible]
;   TYPE: PRT
;   ORGANISM: Ehrlichia [illegible]
US-09-29[illegible]-628-7[illegible]

  Query Match [illegible]
  Best Local Similarity [illegible]
  Matches [illegible]; Conservative [illegible]; Mismatches [illegible]

Qy     74 DMVV[illegible] [illegible]NNWSV[illegible]SAE[illegible]QAYVEN [illegible]SVVALAVVK[illegible]

Db    [illegible] DMVVEKEAHE[illegible] [illegible]VEN[illegible]EYEAVIAE[illegible]

Qy    [illegible]2 S [illegible] KEYA[illegible] [illegible]NKIM [illegible]

Db    1[illegible]4 SYVNK[illegible]NA[illegible] [illegible]ALAKVV[illegible]

Qy    1[illegible]7 KEIKVCVV[illegible] [illegible]MN[illegible] WALI[illegible]W [illegible]

Db    [illegible]14 VEISS[illegible]AV[illegible] [illegible]KEEEE [illegible]YKN[illegible] [illegible]MAI[illegible]LALVIV [illegible]

Qy    [illegible]12 SNENYIENI[illegible]KIKE [illegible]VENYEIA [illegible]KVKNE[illegible]EYVKE[illegible] [illegible]

Db    27[illegible] S[illegible]QEVIVE[illegible]KNSE[illegible]LHE[illegible]EVINEVE[illegible]KNEA[illegible]AYKAY[illegible]EKSEEEIVSEAYIIEIAI [illegible]

Qy    2[illegible] EVLYIKISVS[illegible]S[illegible]SESVEKVYE [illegible]KAHE[illegible]K [illegible]

Db    [illegible] NN[illegible]SKIV[illegible]VNVVEAINE[illegible]KEK[illegible]

RESULT [illegible]
US-09-[illegible]
; Sequence 7[illegible], Application US/[illegible]
; Patent No. [illegible]
; GENERAL INFORMATION:
;   APPLICANT: Reed, Steven G.
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Houghton, Raymond
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE [illegible] AND [illegible]
;   NUMBER OF SEQUENCES: [illegible]
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SEED and BERRY LLP
;     STREET: 6300 Columbia Center, 701 Fifth Avenue
;     CITY: Seattle
;     STATE: Washington
;     COUNTRY: USA
;     ZIP: 9810[illegible]
;   COMPUTER READABLE FORM:
```

```
;          MEDIUM TYPE:  Floppy disk
;          COMPUTER:  IBM PC compatible
;          OPERATING SYSTEM:  PC-DOS/MS-DOS
;          SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;          APPLICATION NUMBER:  US/[illegible]
;          FILING DATE:  2[illegible] JUN 199[illegible]
;          CLASSIFICATION:
;      ATTORNEY/AGENT INFORMATION:
;          NAME:  Maki, David J.
;          REGISTRATION NUMBER:  31,392
;          REFERENCE/DOCKET NUMBER:  [illegible]
;      TELECOMMUNICATION INFORMATION:
;          TELEPHONE:  206-622-4900
;          TELEFAX:  206-682-6031
;   INFORMATION FOR SEQ ID NO:  73:
;      SEQUENCE CHARACTERISTICS:
;          LENGTH:  530 amino acids
;          TYPE:  amino acid
;          STRANDEDNESS:
;          TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-09-106-582-73

  Query Match              5.7%;  Score 97.5;  DB 4;  Length 530;
  Best Local Similarity   22.8%;  Pred. No. [illegible];
  Matches   68;  Conservative   54;  Mismatches  101;  Indels   75;  Gaps   16;

Qy         74 DMVV      [illegible]INNWGVLL[illegible]AFLQAYVKN          -        SVVAFAVVKSE 111
Db        110 DMVVEKFAH[illegible]       V[illegible]KNI[illegible]ANGVF[illegible]YLAVLADSGMIES 163

Qy        112 S       KRYA[illegible]VEVL[illegible]SY[illegible]SAMIMF[illegible]YSGE[illegible]NIKIM 166
Db        164 SYVNKIENAL[illegible] ME[illegible]NEAE         [illegible]YYDALAKDVVLGL-LQHRVADI 213

Qy        167 KEIEV[illegible]VYSL [illegible]    [illegible]MNGMEYAY[illegible]M[illegible]LKFKC-        --WADLIW 211
Db        214 VE[illegible]DAV[illegible]KLFEE     [illegible]YRSA[illegible]IMA[illegible]ALVIV 269

Qy        212 SNPNY[illegible]      [illegible]KSH[illegible]EVENF[illegible]YVEDI      264
Db        270 SDQEVDVE[illegible]KNGELHE[illegible]DVLNLV[illegible]KNEALLAYKAY[illegible]KSF[illegible]VSDAGY[illegible]IEDIAI 329

Qy        265  RVLYDKLSV[illegible]     [illegible]KAHE[illegible]KEV[illegible]KLEEKI[illegible] 316
Db        330 NNI[illegible]VLEV[illegible]VENVVFA[illegible]EV[illegible]VV[illegible]MEV[illegible]KHE[illegible]EKLEEKIS 387

RESULT  7
US-09-066-047-2
; Sequence 2, Application US/09066047A
; Patent No. 6306394
;   GENERAL INFORMATION:
;         APPLICANT: MURPHY, Cheryl
;                    STOREY, James
;                    BELTZ, Gerald A.
;                    COUGHLIN, Richard T.
;         TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF
;                             USE OF GRANULOCYTIC ERHLICHIA
;         NUMBER OF SEQUENCES: 41
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: HALE AND DORR LLP
;              STREET: 60 State Street
;              CITY: Boston
;              STATE: Massachusetts
;              COUNTRY: United States
;              ZIP: 02109
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/[illegible]
;              FILING DATE: 24-Apr-1998
;              CLASSIFICATION: Unknown
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US 60/044,9[illegible]
;              FILING DATE: 25 APR 1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Siperko, Colleen
;              REGISTRATION NUMBER: 39,850
;              REFERENCE/DOCKET NUMBER: 10[illegible]
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (617) 526-6000
;              TELEFAX: (617) 52[illegible]
;   INFORMATION FOR SEQ ID NO: 2:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 546 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         HYPOTHETICAL: NO
;         ANTI-SENSE: NO
;         SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-066-047-2

  Query Match              5.7%;  Score 97.5;  DB 4;  Length 546;
  Best Local Similarity   22.8%;  Pred. No. [illegible];
  Matches   68;  Conservative   54;  Mismatches  101;  Indels   [illegible];  Gaps   [illegible];

Qy         74 DMVV      [illegible]INNWGVLL[illegible]ARLQAYVKN               [illegible]VVA[illegible]VVKSE 111
Db         11 DMVVEKFAH[illegible]       V[illegible]LIKNI[illegible]ANGVF[illegible]YLAVLA[illegible]MIES 64

Qy        112 S----KRYA[illegible]VEVL[illegible]SAMIMF[illegible]NIKIM 166
Db         65 SYVNKIENALF[illegible] ME[illegible]NEAE        [illegible]YYDALAKDVVLGL-LQHRVADI 114

Qy        167 KEIEVSVYSL [illegible]    [illegible]MNGMEYAY[illegible]LKFK[illegible]         WA[illegible] 211
Db        115 VEISSDAVD[illegible]KLFEE     QYKN[illegible]IMA[illegible]AVIV [illegible]

Qy        212 SNPNY[illegible]KE  [illegible]VPNYF[illegible]A   [illegible]SKM[illegible]VENF[illegible]YVE[illegible]   2[illegible]
Db        171 SDQEVDVEIKNSELHE[illegible]DVLNLVF[illegible]KNEAFLAYKAY[illegible]KSF[illegible]VS[illegible]A[illegible] [illegible]

Qy        265  RVLYDKLSVSIDSPIDSESVIKVYF     [illegible]KAHE[illegible]KEVI KLEEKI[illegible] 31[illegible]
Db        231 NNI[illegible]VLEV[illegible]VENVVFA[illegible]VV[illegible]KHE[illegible]LEKLEEKI[illegible] [illegible]

RESULT  8
US-08-975-762-64
; Sequence 64, Application US/08975762
; Patent No. 6207169
;   GENERAL INFORMATION:
;     APPLICANT:  Reed, Steven G
;     APPLICANT:  Lodes, Michael J.
;     APPLICANT:  Houghton, Raymond
;     TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE [illegible]         [illegible]
;     NUMBER OF SEQUENCES:  73
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
```

```
Matches   ;  Conservative   4;  Mismatches   ;  Indels   ;  Gaps   ;

Qy     74 DMVV        NNWSVI   SAFLAYVKN                    VVAPAVVKS III

Db    110 DMVVEKFAHFDIE        VSSNSLRSLIKNIEIELANIVELIRYLAVIAESMIE

Qy        S       KEYAN                   M    K        N        KIM

Db    164 SYVNKIRNA                 M       NEAF                   AKIVVI    HRVA

Qy    167 KEIKVVVY    YE                                  WAIIIW

Db    204 VEISSLAVI  SLISII  KIEII       YENSINIEYEFALAIMAII  IADVIV

Qy    212 SNENYIENISSEIIKI  IVENYEIA     KEMEEKAIEVEKIEIRVENIIYVKII

Db    2   SE  EVIVEIKNSEIHE KEVENIVE  KNEAEIAYEAY  EKIEEEIVIIAIYIIDIAI

Qy    2   EVIVIKISVSI SIIISESVEKVYE             KEKAEIEIKVI KIEEK

Db    330 NNISKIVIIVWVENVVIAINEIEVSIMEEIVVSWEIMKVIEKIIIIKIIIEKIKEKI
```

RESULT 1
US-09-106-182-64
; Sequence 64, Application US/09106182
; Patent No. 6090462
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,182
; FILING DATE: 26-JUN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-106-182-64

```
Query Match          ;  Score   ;  DB   ;  Length   ;
Best Local Similarity   ;  Pred. No.   ;
Matches   ;  Conservative   4;  Mismatches   ;  Indels   ;  Gaps   ;

Qy     74 DMVV        NNWSVI   SAFLAYVKN                    VVAPAVVKS

Db    110 DMVVEKFAHFDIE        VSSNSLRSLIKNIEIELANIVELIRYLAVIAESMIE
```

```
Qy   [illegible]
Db   164 [illegible]ALA[illegible]
Qy   167 KK[illegible]KVSVVS[illegible]KEG  WA[illegible] 2[illegible]
Db   214 VEISSDAV[illegible]SAF[illegible]ADV V 2[illegible]
Qy   2[illegible]2 SNPN[illegible]KSH[illegible]
Db   2[illegible]
Qy   2[illegible]5 [illegible]VFF[illegible]AHL[illegible] 16
Db   [illegible] NN[illegible]VENVVF[illegible]KHL[illegible] [illegible]7
```

RESULT 11
US-[illegible]-4
; Sequence 4, Application US/0900[illegible]
; Patent No. 6013266
; GENERAL INFORMATION:
; APPLICANT: Segers, Ruud P.A.M.
; APPLICANT: Frey, Joachim
; TITLE OF INVENTION: Live attenuated Actinobacillus
; TITLE OF INVENTION: pleuropneumoniae
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo Nobel [illegible] Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: [illegible],570
; FILING DATE: 09-APR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,4[illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7[illegible]
; TELEFAX: (301) 948-[illegible]
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1382 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-[illegible]-4

Query Match 5.5%; Score 94; DB 3; Length 1382;
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches 116; Indels [illegible]4; Gaps [illegible];

```
Qy   [illegible]2 [illegible]NYV[illegible] 75
Db   860 IEALENSIYE[illegible]SRYI[illegible] 91[illegible]
Qy   76 VV[illegible]KKYA[illegible] 1[illegible]
Db   914 NW  AATENL[illegible] 94[illegible]
Qy   1[illegible]2 [illegible]KV[illegible] 16[illegible]
Db   943 EPNNIV[illegible]
```

```
Qy   [illegible]
Db   [illegible]
Qy   [illegible]
Db   [illegible]
Qy   [illegible] 322
Db   [illegible] 1109
```

RESULT 12
US-08-486-099-116
; Sequence 116, Application US/08486099
; Patent No. 6013263
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION ASSOCIATED EVENTS, INCLUDING HEPATITIS
; TITLE OF INVENTION: B VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.[illegible]
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,099
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 43[illegible]
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-03[illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-486-099-116

Query Match 5.4%; Score 9[illegible]; DB [illegible]; Length 607;
Best Local Similarity 21.2%; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

```
Qy   4[illegible]
Db   [illegible] 4[illegible]
Qy   [illegible]
```

[illegible]

```
Qy     [illegible] KLREKISTAE[illegible]NEVEKIESEKE [illegible]
Db     [illegible] KINKE[illegible]VISV[illegible]Y NIV[illegible]KAIKEE [illegible]
```

```
RESULT 14
US-08-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. [illegible]
;  GENERAL INFORMATION:
;    APPLICANT: [illegible]
;    APPLICANT: Matthews, Thomas [illegible]
;    APPLICANT: Wild, [illegible]
;    APPLICANT: [illegible]
;    APPLICANT: [illegible]
;    APPLICANT: [illegible]
;    APPLICANT: Lang[illegible], Alphonse [illegible]
;    TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
;    TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
;    TITLE OF INVENTION: TRANSMISSION
;    NUMBER OF SEQUENCES: [illegible]
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Pennie & Edmonds
;      STREET: 1155 Avenue of the Americas
;      CITY: New York
;      STATE: New York
;      COUNTRY: USA
;      ZIP: [illegible]
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #[illegible], Version [illegible]
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/[illegible]
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: [illegible]
;      FILING DATE: [illegible]
;    ATTORNEY/AGENT INFORMATION:
;      NAME: [illegible]
;      REGISTRATION NUMBER: [illegible]
;      REFERENCE/DOCKET NUMBER: [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (212) [illegible]
;      TELEFAX: (212) 86[illegible]/8864
;      TELEX: 66141 PENNIE
;  INFORMATION FOR SEQ ID NO: [illegible]
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: [illegible] amino acids
;      TYPE: amino acid
```

```
;     STRANDEDNESS:
;     TOPOLOGY: unknown
;   MOLECULE TYPE: protein
US-08-959-597-116

  Query Match             [illegible].4%;  Score 93;  DB [illegible];  Length 607;
  Best Local Similarity   21.2%;  Pred. No. [illegible];
  Matches  [illegible];  Conservative  [illegible];  Mismatches  117;  Indels  86;  Gaps  [illegible];

Qy     4[illegible] [illegible]NNWSV[illegible]AYVKN[illegible]V [illegible]
Db     13 EEAEEEQEENLEAC[illegible] [illegible] 4[illegible]

Qy    1[illegible] VA[illegible]AVVK[illegible] [illegible]PF[illegible] [illegible]
Db     4[illegible] A[illegible]KAM[illegible] [illegible]AVVF[illegible]KN[illegible] [illegible]

Qy    162 N[illegible]VA[illegible]MG[illegible]WA[illegible] 2[illegible]
Db    10[illegible] NIYV[illegible]L [illegible]WV[illegible] 1[illegible]

Qy    2[illegible] [illegible]
Db    152 [illegible]RDI[illegible] 2[illegible]

Qy    259 [illegible]LRE[illegible]AHI[illegible]V [illegible]
Db    211 [illegible]DISLEYRDIISTAEEEDL [illegible]RVHI[illegible]LRKVRAEL[illegible]KE[illegible] 2[illegible]

Qy    [illegible] K[illegible]SEKF [illegible]
Db    261 KLNK[illegible]IV[illegible] N[illegible]KALK[illegible] 2[illegible]


RESULT 15
US-08-475-668A-116
; Sequence 116, Application US/08475668A
; Patent No. [illegible]
;  GENERAL INFORMATION:
;    APPLICANT: Barney, Shawn O.
;    APPLICANT: Lambert, Dennis M.
;    APPLICANT: Petteway, Stephen R.
;    TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
;    TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
;    TITLE OF INVENTION: TRANSMISSION
;    NUMBER OF SEQUENCES: 211
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Pennie & Edmonds LLP
;      STREET: 1155 Avenue of the Americas
;      CITY: New York
;      STATE: New York
;      COUNTRY: USA
;      ZIP: 10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/475,668A
;      FILING DATE: 07-JUN-1995
;      CLASSIFICATION: 424
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Coruzzi, Laura A.
;      REGISTRATION NUMBER: 30,742
;      REFERENCE/DOCKET NUMBER: 7932-026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (212) 790-9090
;      TELEFAX: (212) 869-9741/8864
;      TELEX: 66141 PENNIE
;  INFORMATION FOR SEQ ID NO: 116:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 607 amino acids
;      TYPE: amino acid
;      STRANDEDNESS:
;      TOPOLOGY: unknown
;    MOLECULE TYPE: protein
US-08-475-668A-116

  Query Match             [illegible].4%;  Score 93;  DB [illegible];  Length [illegible];
  Best Local Similarity   21.2%;  Pred. No. [illegible];
  Matches  [illegible];  Conservative  5[illegible];  Mismatches  11[illegible];  Indels  [illegible];  Gaps  [illegible];

Qy     4[illegible] [illegible]AYVKNSV 10[illegible]
Db     1[illegible] [illegible]YYYSGRLSL[illegible] [illegible] 4[illegible]

Qy    1[illegible] [illegible]
Db     4 A[illegible]KAM[illegible]KN[illegible]VN [illegible]K 9[illegible]

Qy    162 N[illegible]WA[illegible]N 2[illegible]
Db    1[illegible] N[illegible]WV[illegible]N 1[illegible]

Qy    21[illegible] [illegible]NI[illegible] 2[illegible]
Db    152 [illegible]AKA[illegible]MHEKKPQSI 2[illegible]

Qy    259 [illegible]VEKVY [illegible]V [illegible]
Db    211 [illegible]RVHI[illegible]RKVRAKL[illegible]KKEA[illegible] 2[illegible]

Qy    [illegible] K[illegible]SEKF 334
Db    261 KLNK[illegible]IV[illegible] N[illegible]KALK[illegible] 294


Search completed: March 19, 2002, [illegible]
Job time: 4617 sec
```

[illegible] Seconds
[illegible] (s)
[illegible] cell updates/sec

[illegible] KRESSERN 41

[illegible]/AA1981.DAT:*
[illegible]/AA1982.DAT:*
[illegible]/AA1983.DAT:*
[illegible]/AA1984.DAT:*
[illegible]/AA1985.DAT:*
[illegible]/AA1986.DAT:*
[illegible]/AA1987.DAT:*
[illegible]/AA1988.DAT:*

[illegible] have a [illegible] being printed. [illegible]

SUMMARIES

	Description
AAY[illegible]	P37/FlaA protein [illegible]
AAY[illegible]	B. burgdorferi [illegible]
AAY[illegible]	B. burgdorferi [illegible]
AA[illegible]	[illegible]
AAW[illegible]	Treponema pallidum [illegible]
AAW[illegible]	A. pleuropneumonia [illegible]
AAW[illegible]	A. pleuropneumonia [illegible]
AA[illegible]	Bacillus subtilis [illegible]
AAW[illegible]	[illegible]
AAY[illegible]	W[illegible] Seq ID N

40	[illegible]	[illegible]	[illegible]	[illegible]	AA[illegible]	[illegible]
41	93	[illegible]	[illegible]	[illegible]	AB[illegible]	Human [illegible]
42	93	[illegible]	124	2[illegible]	AAY2[illegible]	B. burgdorferi [illegible]
43	93	[illegible]	[illegible]	2[illegible]	AAY2[illegible]	B. burgdorferi [illegible]
44	93	[illegible]	6[illegible]	22	AAU[illegible]	Peptide sequence [illegible]
45	93	[illegible]	[illegible]	[illegible]	AAB[illegible]	Human [illegible]

ALIGNMENTS

RESULT 1
AAY2[illegible]
ID AAY2[illegible] standard; [illegible] 41 AA.
XX
AC AAY2[illegible];
XX
DT [illegible]
XX
DE P37/FlaA protein [illegible]
XX
KW P37; FlaA; Lyme disease; [illegible]
XX
OS Synthetic.
OS Borrelia burgdorferi.
XX
FH Key Location/Qualifiers
FT Peptide 1..2[illegible]
FT /label= signal peptide
FT Protein [illegible]..41
FT /label= mature protein
XX
PN W[illegible]
XX
PD [illegible]
XX
PF [illegible]-JAN-[illegible]
XX
PR [illegible]-JAN-[illegible]
XX
PA (INME-) [illegible]
XX
PI [illegible]
XX

```
DR   WPI; 1999-430499/36.
DR   N-PSDB; AAX89529.
XX
PT   Recombinant FlaA/P37 protein antigen, useful in improved detection
PT   of Borrelia burgdorferi
XX
PS   Claim 2; Page 26-28; 34pp; English.
XX
CC   The sequence is the recombinant protein antigen P37/FlaA. FlaA is an
CC   outer sheath protein of the periplasmic flagella of Borrelia burgdorferi.
CC   The P37 protein product can be used as an assay for the detection of
CC   Borrelia burgdorferi infection, the causative agent of Lyme disease. FlaA
CC   is a prominent antigen in the early humoral immune response to Borrelia
CC   burgdorferi infection, and significantly suitable for use in improved
CC   serologic tests for exposure to Lyme disease spirochetes. The detection
CC   of FlaA can augment the set of recombinant molecules that are recognized
CC   early in the course of disease and contribute to improved sensitivity of
CC   early testing for Lyme disease.
XX
SQ   Sequence   341 AA;

  Query Match            100.0%;  Score 1719;  DB 20;  Length 341;
  Best Local Similarity  100.0%;  Pred. No. 8.8e-162;
  Matches  341;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKKKAK[illegible]AELAK[illegible]NYVD 60
Db          1 MKKKAK[illegible]AELAK[illegible]NYVD 60

Qy         61 [illegible]VK[illegible]MVV[illegible]NNW[illegible]SAK[illegible]AYVKNSVVAFAVVKSESKKYACDII 120
Db         61 [illegible]VK[illegible]MVV[illegible]NNW[illegible]SAK[illegible]VVAFAVVK[illegible]KKYAC[illegible] 120

Qy        121 [illegible]VRV[illegible]SAMIM[illegible]SN[illegible]NIKIMKKIEVSVYSLGYEI 180
Db        121 [illegible]VRV[illegible]SAMIM[illegible]SN[illegible]NIKIMKKIEVSVYSLGYEI 180

Qy        181 DLEVI[illegible]MN[illegible]MEYAY[illegible]WAL[illegible]IW[illegible]SRI[illegible]NYFLASSKMR 240
Db        181 [illegible]MN[illegible]MEYAY[illegible]WAL[illegible]IW[illegible]RI[illegible]NYFLASSKMR 240

Qy        241 FFA[illegible]SVI[illegible]K 300
Db        241 FFA[illegible]SVI[illegible]K 300

Qy        301 A[illegible]KN 341
Db        301 AHEIFKKV[illegible]KN 341


RESULT 2
AAY20146
ID   AAY20146 standard; Protein; 339 AA.
XX
AC   AAY20146;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, 107A.aa RB023.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12716.
XX
PR   19-SEP-1997;   97US-0077[illegible].
PR   20-JUN-1997;   97US-0050[illegible].
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (MEDI ) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61843.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 214; 279pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   339 AA;

  Query Match            95.7%;  Score 1645;  DB 20;  Length 339;
  Best Local Similarity  97.6%;  Pred. No. 1.9e-154;
  Matches  327;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          7 S[illegible]LFFLLSIVLFAQEI[illegible]LAEGSKKAEPGELVI[illegible]NYV[illegible]YVYS[illegible]A 66
Db          1 S[illegible]LFFLLSIVLFAQEI[illegible]LAEGSKKAEPGELVI[illegible]NYV[illegible]YVYS[illegible]A 60

Qy         67 [illegible]IVK[illegible]MVV[illegible]INNW[illegible]AKLQAYV[illegible]VRVI 126
Db         61 SGIVKPEDMVVDLGINNWSV[illegible]PSAKLQAYVKN[illegible]VVAFAVVK[illegible]VRVL 120

Qy        127 [illegible]AMIMFF[illegible] 186
Db        121 [illegible]YQ[illegible]AMIMFF[illegible] 180

Qy        187 EDMN[illegible]MF[illegible]WAL[illegible]KAFEV 246
Db        181 EDMNYMF[illegible]WAL[illegible]EMKFKAFEV 240

Qy        247 [illegible]KVKNF[illegible]FYVK[illegible]KAHEIEK 306
Db        241 SKSHSSKE[illegible]FYVK[illegible]KAHAIEK 300

Qy        307 KVLKLKEKISIAEGSEQNEVEKIESEKPEESSEKN 341
Db        301 KVLKLLEKIGM[illegible]KPEESSEKN 335


RESULT 3
AAY20147
ID   AAY20147 standard; Protein; [illegible] AA.
XX
AC   AAY20147;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, 107A.aa RB024.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
```

[illegible] used to develop
[illegible] diseases

[illegible] polypeptides
[illegible] antibodies to members of
[illegible] Lyme disease in
[illegible] products can also

[illegible] Gaps [illegible]

[illegible] MVVLLINNWSV [illegible]

[illegible] MVVLLINNWSV [illegible]

[illegible] AMIMEEKIFFY [illegible]

[illegible] AMIMEEKIFFY [illegible]

[illegible] YAYSMIIKEKGW [illegible]

[illegible] KVEVYLLIYEE [illegible] YAXSMGIIKEKGW [illegible]

[illegible]

[illegible] NEIIYVKIIEV [illegible]

[illegible] SEKISMEEISE_NEV [illegible]

[illegible] AA

[illegible] antibody:

```
XX
FH   Key             Location/Qualifiers
FT   Peptide         [illegible]
FT                   /note= "Signal peptide"
FT   Protein         [illegible]
FT                   /note= "Mature protein"
XX
PN   [illegible] A.
XX
PD   [illegible] MAR [illegible].
XX
PF   14-SEP-[illegible]
XX
PR   [illegible] SEP [illegible]
PR   24-JUL-[illegible]
XX
PA   ([illegible]) [illegible] INT RES [illegible]
XX
PI   Koopman MBH, Kusters [illegible]
XX
DR   WPI; [illegible]
DR   N-PSDB; AA[illegible]
XX
PT   Vaccine to protect pigs against swine dysentery - [illegible]
PT   Treponema hyodysenteriae [illegible]
PT   orally or intranasally
XX
PS   Claim 2; Page 21-22; [illegible]pp; English.
XX
CC   This sequence represents the [illegible]
CC   hyodysenteriae. The endoflagellum consists of at least [illegible]
CC   proteins, this protien forms the sheath of the flagellum and three
CC   proteins, of molecular weights [illegible] make up its core.
CC   Antibodies raised against the sheath protein have been shown to be
CC   bactericidal for T. hyodysenteriae. The [illegible] kD sheath protein can
CC   be used in the production of a vaccine against infections such as
CC   swine dysentery.
XX
SQ   Sequence [illegible] AA;

  Query Match              [illegible]%;  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity    [illegible]%;  Pred. No. [illegible];
  Matches [illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy        [illegible] [illegible]NYV[illegible] [illegible]
Db          4 [illegible]NA[illegible] [illegible]

Qy        [illegible] [illegible]
Db         44 [illegible] [illegible]

Qy        129 [illegible] [illegible]
Db         [illegible] [illegible] [illegible]

Qy        187 [illegible] [illegible]
Db         [illegible] [illegible] [illegible]

Qy        247 SKSHSSKVKN[illegible] [illegible]
Db         [illegible] [illegible] [illegible]

Qy        [illegible] RVLKIKEKI[illegible] [illegible]
Db         [illegible] [illegible] [illegible]


RESULT [illegible]
AAW2[illegible]
ID   AAW2[illegible] standard; [illegible] AA.
```

```
XX
AC   AAW... [illegible]
XX
DT   12 JAN 1998  (first entry)
XX
DE   Treponema pallidum rare outer membrane protein (TROMP 2).
XX
KW   Rare; outer membrane protein; Treponema pallidum; Spirochaetales;
KW   Immune response; Syphilis; TROMP 2; precursor.
XX
OS   Treponema pallidum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..24
FT                   /note= "putative signal peptide"
FT   Cleavage site   22..24
FT                   /note= "cleaves at the C terminal side of Ala 24"
FT   Protein         25..242
FT                   /note= "putative mature protein"
XX
PN   WO9742[illegible] A1.
XX
PD   31-JUL-1997.
XX
PF   23-JAN-1997;   97WO-[illegible].
XX
PR   23-JAN-1996;   96US-0[illegible]948[illegible].
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Blanco DR,  Lovett MA,  Miller JN,  Tempst PJ;
XX
DR   WPI; 1997-393614/36.
DR   N-PSDB; AAT76579.
XX
PT   New isolated Treponema pallidum outer membrane protein - used in the
PT   detection and induction of immune response in an animal to
PT   pathogenic Spirochaetales
XX
PS   Claim 1; Fig 11; 77pp; English.
XX
CC   This rare outer membrane protein (OMP) of species Treponema pallidum,
CC   subspecies pallidum is encoded by a 765 bp sequence (TROMP2 gene).
CC   This is a precursor TROMP2 protein of 242 amino acids. Treponemes
CC   were extracted from infected rabbit testicles and the suspension
CC   was subjected to two low speed centrifugations to remove the
CC   tissue debris.  Treponemes were obtained from the uppermost zone
CC   gradient after two high speed centrifugations involving a density
CC   gradient separation medium. R18 octyl decyl Rhodamine chloride was added
CC   to the treponemal suspension which was incubated at room temperature for
CC   15 minutes and then centrifuged. The treponemal pellet was resuspended
CC   in ice cold sodium citrate buffer, incubated with occasional
CC   vortexing to release the outer membrane from the inner membrane
CC   Treponemal protoplasmic cylinders were removed by centrifugation.
CC   The outer membrane in the supernatant was subjected to SDS-PAGE to
CC   obtain the rare OMP.  The recombinant expression of the rare OMP's
CC   can be used for diagnostic tests to detect syphilis and for
CC   development of host immunity during syphilis.
XX
SQ   Sequence   242 AA;

  Query Match             7.3%;  Score 125.5;  DB 18;  Length 242;
  Best Local Similarity   22.4%;  Pred. No. [illegible];
  Matches   55;  Conservative   31;  Mismatches  134;  Indels   55;  Gaps    8;

Qy         48 PSSIRLDLINYVLTV    TLSAS[illegible]IVKFIDMVVD[illegible]INNW     SVL[illegible]SAKLQAYVKN 100
Db         21 PLAACLSKVNTQAT[illegible]                  [illegible] AWEAAGSKE[illegible]KYFEG 70

Qy        101 SVVAFAVVKSESKRYAGDT[illegible]GVRVLFPSYSQSSAMIMP  -- PEKIFFYSGESGNQFL 155
Db         71 MPQAVRMAGSWQGKF[illegible]EAFF[illegible]VFCKENPQGNNW[illegible]PIEGGS[illegible]YPIPLRCVVSG---- 126
```

```
Qy        [illegible] [illegible]
Db        127           [illegible]VWVWGAGT[illegible]LFALVA[illegible]WKN[illegible] 17[illegible]

Qy        216 Y[illegible]VPNY[illegible]KMEFEAFRV[illegible]KYKN[illegible]YVK[illegible] 2[illegible]
Db        174 HIPQTS  -        RYLGSAQHISFVGFRIRISPSERVDDFY         VPI[illegible]KALA 21[illegible]

Qy        276 [illegible] 280
Db        218 NMHID 222

RESULT 6
AAW81355
ID   AAW81355 standard; Protein; 1657 AA.
XX
AC   AAW81355;
XX
DT   16-APR-1999  (first entry)
XX
DE   A. pleuropneumoniae ApxIV toxin serotype 1 protein sequence.
XX
KW   ApxIV toxin; RTX toxin; attenuated microorganism; subunit vaccine;
KW   A. pleuropneumoniae infection; diagnosis.
XX
OS   Actinobacillus pleuropneumoniae.
XX
PN   EP875574 A2.
XX
PD   04-NOV-1998.
XX
PF   08-APR-1998;   98EP-0201117.
XX
PR   10-APR-1997;   97EP-0201052.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
PI   Frey J,  Segers RPAM;
XX
DR   WPI; 1999-027956/03.
DR   N-PSDB; AAV68471.
XX
PT   Attenuated live Actinobacillus pleuropneumoniae - and a nucleotide
PT   sequence having promoter controlling expression of ApxIV gene, a
PT   subunit vaccine containing pure ApxIV toxin, and an attenuated live
PT   vaccine
XX
PS   Example 1; Page 13-20; 56pp; English.
XX
CC   This sequence represents the ApxIV toxin serotype 1 [illegible]
CC   Actinobacillus pleuropneumoniae strain 4074 (serotype 1 reference
CC   strain). The invention relates to an attenuated live Actinobacillus
CC   pleuropneumoniae microorganism producing no functional ApxIV toxin. The
CC   invention also provides a subunit vaccine for protecting an animal
CC   against infection of A. pleuropneumoniae. The vaccine contains pure ApxIV
CC   toxin and a pharmacologically allowable carrier. The pure ApxIV toxin can
CC   be used for the diagnosis based on the presence of an antibody against
CC   ApxIV [illegible] A. pleuropneumoniae [illegible] A. [illegible]
CC   N.B.  This sequence was indexed from EP875574, which is the first [illegible]
CC   country equivalent to JP10290695.
XX
SQ   Sequence   1657 AA;

  Query Match             6.4%;  Score 11[illegible];  DB 20;  Length 1657;
  Best Local Similarity   21.5%;  Pred. No. 0.45;
  Matches   68;  Conservative   51;  Mismatches  1[illegible];  Indels   [illegible];  Gaps   1[illegible];

Qy         22 IDGLAEGSKEAEIGELVLDFAK    LAKDFSSIK    L[illegible]INYVLTVT[illegible]AS[illegible]IVKFELM [illegible]
Db        860 EALFNSIFKQSPFNALYDLSEYLSEINPPIEWK[illegible]SFYIE   YAKAQGFYE      91[illegible]
```

```
[illegible] VVA [illegible] LLVEVLFFSYS [illegible]
[illegible] TVYEDT 999
[illegible] KMEEKAFFVSKS [illegible]
[illegible] MEEYHE [illegible]
[illegible] ESIEKIKAHEIE 305
[illegible] KHAIW [illegible]

[illegible] AA

[illegible] ApxIV [illegible]
[illegible] ApxIV [illegible] attenuated [illegible]

[illegible] ApxIV [illegible]
[illegible] Actinobacillus
[illegible] ApxIV toxin. The
[illegible] ApxIV
[illegible] ApxIV [illegible]
[illegible] antibody against
[illegible] from A. suis
[illegible] is the first major
```

```
Best Local Similarity  21.7%;  Pred. No. [illegible]
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy     22 [illegible]HAESKEAE[illegible]VIIFAE[illegible]
Db   1008 [illegible]YAKA[illegible]
Qy     76 VVI[illegible]NNW[illegible]AVYKNSVVA[illegible]AVVKS[illegible]
Db   1052 [illegible]NW[illegible]AAISN[illegible]
Qy    134 [illegible]SSAMIMEIEKI[illegible]YS[illegible]
Db   1091 EKNNIL[illegible]
Qy    196 N[illegible]
Db   1148 NNENEAEL[illegible]VNY AEVKEEEV NE[illegible] MEEYHE[illegible]
Qy    250 HSSKVKNE[illegible]YVKE[illegible]ESIEKIKAHEIE [illegible]
Db   1197 [illegible]SVIVKSEYSHV[illegible]KHAIW [illegible]
Qy    306 KEVLKIEEKISIAE[illegible]SE [illegible]
Db   1247 [illegible]SILEEGK[illegible]N[illegible]Y [illegible]

RESULT 6
AAR34719
ID   AAR34719 standard; Protein; 444 AA.
XX
AC   AAR34719;
XX
DT   17-AUG-1993 (first entry)
XX
DE   Bacillus subtilis srfA operon ORF [illegible]
XX
KW   Multienzyme complex surfactin synthetase; MCSS; [illegible]
XX
OS   Bacillus subtilis.
XX
PN   EP540074-A.
XX
PD   05-MAY-1993.
XX
PF   [illegible]
XX
PR   [illegible]
PR   02-SEP-1992; [illegible]
XX
PA   (ENIE ) ENIRICERCHE SPA.
XX
PI   [illegible]
PI   Rodriguez [illegible]
XX
DR   WPI; 1993-145417/18.
DR   N-PSDB; AAQ4[illegible]
XX
PT   Multienzyme complex surfactin synthetase DNA [illegible]
PT   Bacillus subtilis, and used [illegible]
PT   pharmaceutical prods.
XX
PS   [illegible]
XX
CC   The Bacillus subtilis chromosomal DNA [illegible] srfA
CC   operon which encodes the multienzyme complex surfactin synthetase
CC   (MCSS). Analysis of the sequence showed [illegible] potentially
CC   coding for proteins [illegible]
CC   srfA operon promoter and a presumed terminator [illegible]
CC   [illegible] the stop codon of the [illegible] DNA [illegible]
CC   the ORF4 stop [illegible]
CC   potentially coding for different [illegible]
```

```
CC   and are potentially transcribed in the opposite direction to strA from
CC   a single promoter situated between ORF7 and ORF8.  ORF8 is
CC   transcribed in the same direction as the strA operon.  A further
CC   two potential coding sequences were identified in the region immediately
CC   downstream of ORF4, namely ORFX and ORFY.
CC   See also AAR34712-21.
XX
SQ   Sequence   444 AA;

  Query Match             [illegible];  Score 109;  DB 14;  Length 444;
  Best Local Similarity   21.6%;  Pred. No. [illegible];
  Matches   75;  Conservative   42;  Mismatches  126;  Indels  104;  Gaps   [illegible];

Qy         29 SKRAEPGELVLDFAELARDPSSIRLD[illegible]INYVD [illegible] YVYSGASGI  69
Db         [illegible] SKS[illegible] [illegible]AISAED[illegible] [illegible]IYKN[illegible] 125

Qy         70 [illegible] VKFEDMVVL[illegible]NNWGVLL [illegible] EGARLQAYVKN 100
Db        126 IRV[illegible]PKLLAI[illegible]VIA[illegible]KEK[illegible]MVEQLE 185

Qy        1[illegible] SVVA[illegible]AVVE[illegible] [illegible] 160
Db        186 ILGIPAIGVE[illegible]NE [illegible] 2[illegible]3

Qy        161 DNIE[illegible]MN[illegible]EK[illegible]NYIPNI 220
Db        234 EEDK[illegible] [illegible]ENKKAD [illegible]TAYDL 274

Qy        221 SSRI[illegible]KDDVPNV[illegible]FVEDL[illegible]KLSVSIDSDID 280
Db        275 SSHV[illegible] [illegible]YEKLN [illegible]SMI 321

Qy        281 SESVEKVYEISGIESLEKLKAHEIEK PVLKLEEKI     SIAEGSF 322
Db        322 SQA[illegible]GCRF 368

RESULT [illegible]
AAW81168
ID   AAW81168 standard; Protein; 947 AA.
XX
AC   AAW81168;
XX
DT   0[illegible] MAR 1[illegible]  (first entry)
XX
DE   Transcriptional regulatory factor RING3.
XX
KW   Human; transcriptional regulatory factor; RING3; TSB; cancer;
KW   testis specific bromodomain; testicular cell proliferation.
XX
OS   Homo sapiens.
XX
PN   WO9848[illegible]15-A1.
XX
PD   2[illegible]-OCT-1998.
XX
PF   17-APR-1998;   98WO-JP01782.
XX
PR   18-APR-1997;   97JP-0116402.
XX
PA   (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI   Jones MH;
XX
DR   WPI; 1998-[illegible];
DR   N-PSDB; AAV68343.
XX
PT   Transcriptional regulator gene containing bromodomain sequence - may
PT   be expressed in testis tissue and is useful in treatment of cancer
PT   and other proliferative disorders
XX
PS   Claim 1; Page 19-24; 42pp; Japanese.
XX
CC   The present sequence represents the human transcriptional regulatory
CC   factor RING3, which is isolated from testicular cells. RING3 contains
CC   a testis specific bromodomain (TSB) which is expressed specifically
CC   in testis tissue and also expressed in certain [illegible]. The
CC   transgenic cells may be used to express RING3 which is a TSB expressing
CC   protein. The TSB expression product can be used in the treatment of
CC   cancer and other proliferative disorders, and in screening of compounds
CC   for ability to bind to it (e.g. for use as drugs to modulate [illegible]
CC   transcriptional regulation). DNA capable of hybridising to RING3
CC   polynucleotides may be used for construction of probes and primers.
XX
SQ   Sequence   947 AA;

  Query Match             [illegible].2%;  Score [illegible];  DB [illegible];  Length 947;
  Best Local Similarity   18.[illegible]%;  Pred. No. [illegible];
  Matches   [illegible];  Conservative   80;  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy          1 MKRKAKS[illegible]VLFAQ[illegible]SKRA[illegible]ALLAR[illegible]SSI[illegible]INYV 5[illegible]
Db        118 M[illegible]NIAVSSAKEK[illegible]EKVEK[illegible] 1[illegible]

Qy         60 DYVY[illegible]VKI[illegible]DMVV[illegible]NNW[illegible]AIVK     NCVVALAVVK[illegible]SKEY 1[illegible]
Db        178 PSVEPKIS  [illegible]PENVV[illegible]AIVKI[illegible]QIAAQVIE[illegible]AISAVKASSI 2[illegible]

Qy        116 A[illegible]DIIDGVKVLFPSYSQSSAMIMEFFK       [illegible]NIEIMK[illegible] 1[illegible]
Db        231 --PSPIFTEKS VALPPIKENMPKNVLP   [illegible]Y    NVVEIVKVIE[illegible]

Qy        169 IK  ---   --    VSVYSE[illegible]   YEI   [illegible]EVIEEIMNEMEY 1[illegible]
Db        27[illegible] [illegible]KENMLAKKH[illegible]VNAL[illegible]VVKNEM[illegible]KEKM[illegible]NQ[illegible]YK [illegible]

Qy        195 -AYSMCILEPK[illegible]WAPLIWSN      PNYIPNI[illegible]VENYP[illegible]KMPEKAPPVS 24[illegible]
Db        3[illegible] [illegible]AYS     PAAEVKLMPMN[illegible]YKYNE[illegible]HEVVIMAEM[illegible]      VE 3[illegible]

Qy        24[illegible] [illegible]VENF[illegible]VKI[illegible]EL[illegible]EKEAH       [illegible]
Db        370 ETHESKIP--IEPVESMPLCYIKTD[illegible]        [illegible]ENINEAS[illegible]NSSD 41[illegible]

Qy        303      EIEKRVLKLEEKIS        IAEGSF[illegible]NEVEKIE[illegible]EKIE[illegible]SSEK 34[illegible]
Db        414 [illegible]SEEERVKRLAK[illegible]LKAVH[illegible]VE[illegible]NKKKEKSKKEKKKEK 4[illegible]

RESULT 1[illegible]
AAY07114
ID   AAY07114 standard; Protein; 947 AA.
XX
AC   AAY07114;
XX
DT   02-JUL-1999  (first entry)
XX
DE   WO9904265 Seq ID No 385.
XX
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; [illegible] cancer;
KW   prostate cancer.
XX
OS   Homo sapiens.
XX
PN   WO9904265-A2.
XX
PD   28-JAN-1999.
XX
PF   15-JUL-1998;   98WO-US14679.
XX
PR   22-JUN-1998;   98US-[illegible]
PR   17-JUL-1997;   97US-0061[illegible]4.
PR   10-OCT-1997;   97US-0061[illegible]
PR   10-OCT-1997;   97US-0061765.
```

[illegible]

```
DT   [illegible]-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein [illegible]
XX
KW   Antigenic protein; vaccine; Lyme disease; [illegible]
XX
OS   Borrelia burgdorferi.
XX
PN   WO99[illegible]-A1.
XX
PD   [illegible]-FEB-1999.
XX
PF   18-JUN-1998;   98WO-US[illegible]
XX
PR   03-SEP-1997;   97US-[illegible]
PR   2[illegible]-JUN-1997;   97US-[illegible]
PR   22-JUL-1997;   97US-[illegible]44
PR   2[illegible]-JUL-1997;   97US-[illegible]
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   [illegible]
XX
DR   WPI; 1999-[illegible]/[illegible].
DR   N-PSDB; AAX[illegible].
XX
PT   New isolated Borrelia burgdorferi [illegible] used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim [illegible]; Page [illegible]; [illegible]pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for prevention or attenuation of
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus
XX
SQ   Sequence   1087 AA;

  Query Match            [illegible]  Score [illegible]4[illegible]  [illegible]
  Best Local Similarity  2[illegible]  [illegible]
  Matches [illegible]  Conservative  44;  Mismatches [illegible]  Indels [illegible]  Gaps [illegible]

[illegible]

RESULT [illegible]2
```

```
AAY[illegible]
ID   AAY[illegible] standard; Protein; 1119 AA.
XX
AC   AAY[illegible];
XX
DT   [illegible] 1999 (first entry)
XX
DE   B. burgdorferi antigenic protein, 1742.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071 A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   [illegible]-SEP-1997;   97US-0057483.
PR   26-JUN-1997;   97US-0050[illegible].
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (MEDI ) MEDIMMUNE INC.
XX
PI   Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR   WPI; 1999-185[illegible]/16.
DR   N-PSDB; AAX61631.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 124-125; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   1119 AA;

  Query Match             6.1%;  Score 154.5;  DB 20;  Length 1119;
  Best Local Similarity   21.7%;  Pred. No. 0.86;
  Matches   [illegible];  Conservative   44;  Mismatches  110;  Indels   [illegible];  Gaps   [illegible];

Qy         [illegible] VEN[illegible]VVA[illegible]AVVE[illegible] ... AMIM          [illegible]EK[illegible]FY 11[illegible]
Db          46 VEN[illegible] ... [illegible]KAAI ... VAK[illegible]AMLTN[illegible]LAHIAN[illegible]FEGKE[illegible] 96

Qy         14[illegible] [illegible]         HA[illegible]IEVFVYSL[illegible]Y EIELEVLFEDMNCM 192
Db          97 [illegible]INKNVNE[illegible]IAPNNKYIE[illegible]ITENYGLNNLVG[illegible] 15[illegible]

Qy         [illegible] EYA[illegible] ... [illegible]PLA[illegible] 2[illegible]
Db         156 EEKESEIEEEKE ... PNINEEEE ... SVN ... NE[illegible] 20[illegible]

Qy         25[illegible] [illegible] 2[illegible]
Db         20[illegible] [illegible]NSKYSNNNN[illegible] 2[illegible]

Qy         [illegible]   EAEIEEEVIE   [illegible]A[illegible] [illegible]
Db         2[illegible] [illegible] SKNEI[illegible]EV [illegible]

Qy         [illegible] VEEIE[illegible]
```

```
                 [illegible]
Db        327 N[illegible]ENSKNKEAE 3[illegible]


RESULT 13
AAG82627
ID   AAG82627 standard; Protein; [illegible] AA.
XX
AC   AAG82627;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis open reading frame protein sequence SEQ ID NO:[illegible].
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW   vaccination; endocarditis.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO200134809 A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30782.
XX
PR   09-NOV-1999;   99US-0164258.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
DR   N-PSDB; AAH53477.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis
XX
PS   Claim 18; Page 621; 2188pp; English.
XX
CC   AAH52004 to AAH53670 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC   (I) and (II) can have antibacterial activity and therefore can be used
CC   in vaccination. The nucleic acids (I) may be used to produce the
CC   S. epidermidis polypeptides (II) via the production of vectors
CC   containing them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the infection.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH[illegible] to
CC   AAH[illegible] represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH[illegible] to
CC   AAH[illegible]98 represent oligonucleotide sequences and primers which are used
CC   in the exemplification of the present invention.
CC   N.B. The present invention specifically claims all the polynucleotide
CC   sequences given in the sequence listing of the present specification,
CC   however the sequence listing only goes up to SEQ ID NO:4454 so even
CC   though sequences are given in the disclosure for SEQ ID NO:445[illegible] to 44[illegible],
CC   no sequences are present for SEQ ID NO:445[illegible] to 44[illegible].
XX
SQ   Sequence   [illegible] AA;

  Query Match             [illegible]%;  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   21.7%;  Pred. No. [illegible];
  Matches   [illegible];  Conservative   [illegible];  Mismatches   [illegible];  Indels   [illegible];  Gaps   [illegible];

Qy        [illegible] [illegible]VEVNVYSL[illegible] 1[illegible]
Db          1 [illegible]
Qy        1[illegible] [illegible]EIE[illegible]   VLFEDMNCMEEVEM[illegible]   [illegible]EEEVAE[illegible] 2[illegible]
Db        [illegible] [illegible] 14[illegible]
```

```
                                      EIKAEEVSK [illegible]
[illegible] EIAIBEEK[illegible]IKI [illegible]
[illegible] [illegible]
[illegible] NAKIVA[illegible] EAVINKEMI AI [illegible]
[illegible] NEVIE [illegible]
[illegible] NBYIE[illegible]

[illegible] AA

[illegible] bacterial infection
[illegible] epidermidis
[illegible] and treating bacterial infections
[illegible] epidermidis open reading
[illegible] sequences
[illegible] sequences have
[illegible] The sequences
[illegible] bacterial infections.
[illegible] can be used to
[illegible] epidermidis [illegible]
[illegible] of the printed
[illegible] directly from the

[illegible] length [illegible]
[illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]
[illegible] NIKIMKPIKVSVYSL 176
[illegible] APKQAIVNEYQR [illegible]
[illegible] 208
```

```
Db      [illegible] HYIVEIDKEIEV[illegible]KN[illegible] VAI[illegible]VVNE[illegible]VMIAGAKI [illegible]
Qy      [illegible] LIWENENYIEN [illegible] VENYI[illegible]EM          EIKAEEVSK [illegible]
Db      [illegible] EIKISEENYIENWNI[illegible]AKVIEKIKI[illegible]YENN [illegible]EIAIBEEK[illegible]IKI [illegible]
Qy      [illegible] SBSSKVKNE[illegible]VKI[illegible]
Db      [illegible] VHIIAYQA[illegible]AKNE[illegible]AGKNAKIVA[illegible] EAVINKEMI AI [illegible]
Qy      [illegible] EKIKAHE[illegible]KEVIKIEEKISIAE[illegible]NEVIE [illegible]
Db      [illegible] KKYQ[illegible]HINAGME[illegible]AI[illegible]AAIYAINBYIE[illegible]

RESULT [illegible]
AAW[illegible]
ID   AAW[illegible] standard; [illegible] AA.
XX
AC   AAW[illegible];
XX
DT   [illegible] MAR [illegible] (first entry)
XX
DE   H. pylori [illegible] protein.
XX
KW   GHP[illegible] protein; Helicobacter infection; [illegible]
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   W[illegible]-A1.
XX
PD   [illegible]
XX
PF   [illegible] APR [illegible]; [illegible]
XX
PR   24 [illegible]
PR   [illegible] APR [illegible]
PR   24 JUN 1997;   [illegible]US [illegible]
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   ([illegible]NME-) MERIEUX [illegible] PASTEUR MERIEUX [illegible]
XX
PI   A. Garawi A., Kleanthous H., Miller [illegible]
XX
DR   WPI; 1998-[illegible]/4[illegible].
DR   N-PSDB; AAX144[illegible].
XX
PT   New isolated Helicobacter polynucleotides [illegible] products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastro-intestinal diseases
XX
PS   Claim [illegible]; Page [illegible]; [illegible]pp; English.
XX
CC   This sequence represents a Helicobacter pylori [illegible] of the
CC   invention. The polypeptides can be used for prevention or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic and atrophic gastritis and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence   45[illegible] AA;

  Query Match          [illegible]%;  Score [illegible];  DB [illegible];  Length 45[illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches  [illegible];  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy      [illegible] QAYVKNGVVA[illegible]AVVK[illegible]KE  YA[illegible]VEVI[illegible]AMIM  [illegible]IKI[illegible] [illegible]
Db      [illegible] QAEKKNBKI[illegible]NESISKBV[illegible]K[illegible]N[illegible]AIKI[illegible]K[illegible]IKIKM[illegible]YNIENENE[illegible]
```

```
Qy    148 [illegible] MNGM[illegible] 207
Db    131 GDDESQ       SNEEVLKILLYSFEKL[illegible]Y    IKNL[illegible]   --   -   [illegible]GSKCHSGWL 168

Qy    208 [illegible]      V[illegible]NY[illegible]KM    -    [illegible]KAF    KV[illegible] 247
Db    169 SEMY [illegible] CIE[illegible] 224

Qy    248 KSHSSKV   - KN[illegible]YVK[illegible]VF KVYE[illegible]ESLKKLKAHE 303
Db    225 PSELSKVAV[illegible]YAK[illegible]K    PI[illegible]NLGKR        27[illegible]

Qy    304 IFKEVLKLKLKI[illegible] [illegible]QN[illegible]VLKIE 330
Db    274 YFKK[illegible]VK[illegible]NE[illegible]QNKIMTIE 301
```

Search completed: November 4, 2002, 22:57:54
Job time : 85 secs

[illegible]

[illegible] VIEA[illegible] [illegible]SEKPEESSPKN 341

[illegible]

[illegible] being printed.

[illegible]

Description

[illegible] borrelia af
[illegible] borrelia af
[illegible] spirochaeta
[illegible] spirochete
[illegible] spirochete
[illegible] treponema
[illegible] treponema p
[illegible] treponema p
[illegible] treponema p
[illegible] zarobacteri
[illegible] haemophilus
[illegible] arabidopsis
[illegible] oryctolagus
[illegible] thermoanaer
[illegible] arabidopsis

[illegible]

ALIGNMENT

```
RESULT 1
[illegible]7Y[illegible]
ID   [illegible]7Y[illegible]    PRELIMINARY;    PRT;   [illegible] AA.
AC   [illegible]7Y[illegible];
DT   01-DEC-2001 (TrEMBLrel. [illegible], Created)
DT   01-DEC-2001 (TrEMBLrel. [illegible], Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. [illegible], Last annotation update)
DE   Flagellin A.
GN   FLAA.
OS   Borrelia afzelii.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=29518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN A91;
RX   PubMed [illegible];
RA   Panelius J., Lahdenne P., Saxen H., Heikkila T., Seppala I.;
RT   "Recombinant Flagellin A [illegible]
RT   Stricto, B. afzelii, and B. garinii in Serodiagnosis of Lyme
RT   Borreliosis.";
RL   J. Clin. Microbiol. [illegible]
DR   EMBL; AY[illegible]; AAL[illegible].1; -.
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             [illegible]
  Best Local Similarity   [illegible]4.[illegible]%;  Pred. No. [illegible]
  Matches [illegible]2[illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy          1 MKKKAKS[illegible]SIV[illegible]A[illegible]INYV[illegible]
Db          1 MKKKAKS[illegible]SIA[illegible]INYV[illegible]

Qy         61 YVYSGAS[illegible]IVK[illegible]MVV[illegible]NNWSVL[illegible]AYVKNSVVA[illegible]AVVKSESKKYA[illegible]
Db         61 YVYSGAS[illegible]IVK[illegible]MVV[illegible]NWSVL[illegible]AYVKNSVVA[illegible]VKSESKKYA[illegible]

Qy        121 [illegible]VKVL[illegible]AMIM[illegible]K[illegible]IN[illegible]KV[illegible]VYS[illegible]
Db        121 [illegible]VKVL[illegible]SY[illegible]AMIM[illegible]K[illegible]YS[illegible]NIK[illegible]KV[illegible]VYS[illegible]YEV[illegible]
```

```
Qy      181 DLEVLFEDMNGMEYA[illegible]WA[illegible]I[illegible]SLA[illegible]KMS 2[illegible]
            [illegible]
Db      181 DLEVLFEDMNGMEYAY[illegible]WA[illegible]WANENYIENLARTIELIV[illegible]LA[illegible]KMS 24[illegible]

Qy      241 [illegible]SV[illegible]EEKLE 300
            [illegible]
Db      241 EEAEKVSK[illegible]SV[illegible]K[illegible]K 300

Qy      301 AHEIEKKVLE[illegible]N 341
            [illegible]
Db      301 AHEIEKKV[illegible]N 341


RESULT 2
Q93Y4
ID   Q93Y4        PRELIMINARY;      PRT;   342 AA.
AC   Q93Y4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Flagellin A.
GN   FLAA.
OS   Borrelia garinii.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=29519;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=46;
RX   MEDLINE=[illegible];
RA   Panelius J., Lahdenne P., Saxen H., Heikkila T., Seppala I.J.;
RT   "Recombinant Flagellin A Proteins from Borrelia burgdorferi Sensu
RT   Stricto, B. afzelii, and B. garinii in Serodiagnosis of Lyme
RT   Borreliosis.";
RL   J. Clin. Microbiol. 39:4013-4019(2001).
DR   EMBL; AY[illegible]
SQ   SEQUENCE   342 AA;  38559 MW;  5AD603AB0958459D CRC64;

  Query Match             95.4%;  Score 1640.5;  DB 2;  Length 342;
  Best Local Similarity   95.[illegible]%;  Pred. No. [illegible];
  Matches  326;  Conservative    9;  Mismatches    6;  Indels    1;  Gaps    1;

Qy        1 MKKKAKSILFLLSIVLEAQEIDGLAKGSKKAEPGELVLDFAELARGESSTRLDLTNYVL 60
            [illegible]
Db        1 MKKKAK[illegible]FAELA[illegible]NYVL 60

Qy       61 YVY[illegible]SVVA[illegible]VV[illegible] 120
            [illegible]
Db       61 YVY[illegible]SVVA[illegible]VV[illegible] 120

Qy      121 LGVPVLEP[illegible]DN[illegible] 180
            [illegible]
Db      121 LGVPVLEP[illegible]NN[illegible] 180

Qy      181 DLEVLFEDMNGMEYA[illegible]NSS[illegible]KMS 240
            [illegible]
Db      181 DLEVLFEDMNGM[illegible]NSS[illegible]KMS 240

Qy      241 EEA[illegible]SE[illegible]EEKLE 300
            [illegible]
Db      241 EEAEKVSK[illegible]SE[illegible]EEKLE 300

Qy      301 AHEIEKKVLEKEKKISIAE[illegible]KPEKSPK 349
            [illegible]
Db      301 AHEIEKKV[illegible]K 341


RESULT 3
Q9R5Z8
ID   Q9R5Z8       PRELIMINARY;      PRT;   337 AA.
AC   Q9R5Z8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   FLAA FLAGELLAR filament outer layer polypeptide.
OS   Spirochaeta aurantia.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
OX   NCBI_TaxID=147;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93202439; PubMed=8454190;
RA   Parales J.Jr., Greenberg E.P.;
RT   "Analysis of the Spirochaeta aurantia flaA gene and transcript.";
RL   FEMS Microbiol. Lett. 106:245-251(1993).
SQ   [illegible]

  Query Match             23.4%;  Score 4[illegible];  DB 2;  Length 337;
  Best Local Similarity   30.4%;  Pred. No. 2.4e-21;
  Matches  100;  Conservative   [illegible];  Mismatches  124;  Indels   [illegible];  Gaps   [illegible];

Qy        1 MKKKAKSILFLLSIVLEAQ[illegible]NYVL 6[illegible]
            [illegible]
Db        1 MKK       FEAIGAALFVGNSGAFAEQA        [illegible] 4[illegible]

Qy       61 YVYS[illegible]SKKYA[illegible]
            [illegible]
Db       48 YSKQAGSA[illegible]

Qy      118 [illegible]NIKKMK 1[illegible]
            [illegible]
Db      107 [illegible]

Qy      168 [illegible]
            [illegible]
Db      167 [illegible]

Qy      229 [illegible] 294
            [illegible]
Db      2[illegible] [illegible] 29[illegible]

Qy      289     EKVVEISGIESLKKIKAHEIEK 3[illegible]
            [illegible]
Db      29[illegible] [illegible] 314


RESULT 4
Q9LAI7
ID   Q9LAI7       PRELIMINARY;      PRT;   [illegible] AA.
AC   Q9LAI7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   FlaA 1 (Fragment).
GN   FLAA 1.
OS   spirochete str. PERS A.
OC   Bacteria; Spirochaetales; Spirochaetaceae.
OX   NCBI_TaxID=120784;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=[illegible];
RA   Sato T., Kuramitsu H.K.;
RT   "[illegible]
RT   spirochetes in advanced periodontal pockets.";
RL   Arch. Oral Biol. 0:0-0(2000).
DR   EMBL; AF132816; AAF64553.1; -.
FT   NON_TER       1      1
FT   NON_TER     285    285
SQ   [illegible]

  Query Match             20.0%;  Score 34[illegible];  DB 2;  Length 28[illegible];
  Best Local Similarity   29.4%;  Pred. No. 4.2e-1[illegible];
  Matches   90;  Conservative   56;  Mismatches  114;  Indels   4[illegible];  Gaps    8;

Qy       1[illegible] LSIVLEAQEIDGLAKGSKKAEPGELVLDFAELARGESSTRLDLTNYVLYVYS[illegible]
            [illegible]
Db        2 LNAGIMADKGGMTHNKK       TVLDYASLA[illegible]       ESYT[illegible]       EGK [illegible]
```

```
[illegible] LAVVKN[illegible]VVA[illegible] [illegible]VEVI[illegible]FSY [illegible]
[illegible]N[illegible]VAHAA[illegible] [illegible]VEVI[illegible]FIWI[illegible]

[illegible]ES[illegible]N [illegible]NIKIMKEIKV [illegible]
[illegible]NV[illegible]AI [illegible]VVKNV[illegible]VIKSIAV [illegible]

[illegible]NIESEIIKIIVEN [illegible]
[illegible]VESEEVE[illegible]IYEV [illegible]

[illegible]SVEKVYEI[illegible]
[illegible]FALEDIWGLAEE 2[illegible]
```

[illegible]

```
[illegible]NSAENEVA [illegible]
[illegible]IYEV [illegible]
[illegible]AYYVMA[illegible]
[illegible]NIKIMKE[illegible]ENMNIMEYA [illegible]
[illegible]VVKNV[illegible]YVMMEI[illegible]EVHE [illegible]
[illegible]NIESEIIKIIVEN[illegible]KAEEVSKEHSEKVK 25[illegible]
[illegible]VESEEVE[illegible]EMVIEDAAHAGG 240
[illegible]SVEKVYE[illegible]
[illegible]FALEDIWGLAEE[illegible]
```

```
ID   [illegible]AI8      PRELIMINARY;      PRT;   2[illegible] AA.
AC   [illegible]AI8;
DT   01-OCT-200[illegible] (TrEMBLrel. [illegible], Created)
DT   01-OCT-200[illegible] (TrEMBLrel. [illegible], Last sequence update)
DT   01-OCT-200[illegible] (TrEMBLrel. [illegible], Last annotation update)
DE   FlaA [illegible] (Fragment).
GN   FLAA [illegible].
OS   Treponema [illegible].
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=[illegible];
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A[illegible];
RA   Sato T., Kuramitsu H.K.;
RT   "Polymerase chain reaction [illegible] flaA [illegible] genes [illegible]
RT   spirochetes in advanced periodontal pockets.";
RL   Arch. Oral Biol. [illegible]
DR   EMBL; AF[illegible]; AAF64[illegible];
FT   NON_TER     [illegible]
FT   NON_TER     28[illegible]    28[illegible]
SQ   SEQUENCE   28[illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score [illegible];  DB [illegible];  Length 28[illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches   89;  Conservative   [illegible];  Mismatches  [illegible];  Indels   4[illegible];  Gaps   [illegible];

Qy         1[illegible] LSIVLFAQ[illegible]LA[illegible]SKPAE[illegible]VI[illegible]AF[illegible]AE[illegible]INYV[illegible]VY[illegible]A[illegible]VKI [illegible]

Db          2 LNADIMADKGGMIHNEE       VIIYAGLAI          [illegible]Y[illegible]          [illegible]K [illegible]

Qy         7[illegible] EDMVV[illegible]NNW[illegible]VL[illegible]AFL[illegible]AYVKN[illegible]VVAFAVVK[illegible]YA[illegible]VEVI[illegible]FSY [illegible]

Db         [illegible]8 AIMES[illegible]LAVA[illegible]WLVVIN[illegible]AENIVAHAA[illegible]VI[illegible]AFV[illegible]AK[illegible]A[illegible]VEVIFFIWI[illegible]

Qy        1[illegible]2 QSSAMIMPPPKIPPYS       [illegible]ES[illegible]N              [illegible]NIKIMKEIKV [illegible]

Db         98 NANAMIKEA[illegible]VI[illegible]AYYVMA[illegible]NV[illegible]KA[illegible]VVKNV[illegible]VIK[illegible]IAV [illegible]

Qy        172 SVYS[illegible]VLIF[illegible]MNIM[illegible]YAY[illegible]M[illegible]KEK[illegible]WA[illegible]NIESEIIKIIVEN [illegible]

Db        15[illegible] NIY[illegible]MNYPH LYVMMEI[illegible]EVHEYIM[illegible]WKI[illegible]VWNN[illegible]VE[illegible]EVE[illegible]IYEV [illegible]

Qy        2[illegible] YEIAS[illegible]KMEIKA[illegible]VSK[illegible]SKVKN[illegible]YVK[illegible]YIK[illegible]SVEKVYE[illegible]

Db        2[illegible]7 YIASIIHVVEF[illegible]MVIE[illegible]AAHA[illegible]YV[illegible]YIK[illegible]YIKAV[illegible]AL[illegible]IW[illegible]LAEE [illegible]

Qy        2[illegible]2 [illegible]ESIE 2[illegible]

Db        2[illegible]7 [illegible]ALEKE 2[illegible]


RESULT 7
[illegible]
ID   [illegible]      PRELIMINARY;      PRT;   2[illegible] AA.
AC   [illegible];
DT   01-[illegible] (TrEMBLrel. [illegible], Created)
DT   01-[illegible] (TrEMBLrel. [illegible], Last sequence update)
DT   01-[illegible] (TrEMBLrel. [illegible], Last annotation update)
DE   FlaA homolog [illegible]
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=[illegible];
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Porcella S.F., [illegible], Norris S.J.[illegible];
RT   "Treponema pallidum FlaA homologues.";
RL   Submitted (AUG-[illegible]) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; [illegible]; AA[illegible];
SQ   SEQUENCE   24[illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score [illegible];  DB [illegible];  Length 24[illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches   [illegible];  Conservative   [illegible];  Mismatches  [illegible];  Indels   [illegible];  Gaps   [illegible];
```

```
Qy       48 PSSIRLDLINYVLYV    YRQAIIVKIEDMVVLIINNW  - SVLLIISARLQAYVKN 100
Db       21 PLAAQRFVNYQAYFIIIIGAS           IIQRII AWRAAGSRIIIIKGIIIAYFFI 70

Qy      101 SVVAIAVVEILSEFYAIIIIIIVEVIIIFYRQIAAMIM     IIKIIFYYIHIRIRQPL 155
Db       71 MPQAVRMAGSWQGKLEEAFFIVEIFKENRQGNNWLIIIPIKGGSDYEIIPLRGVVII       126

Qy      156 GKGLIINIKIMKRIKVIVIIIIIIIIIIIVIIIIMNIMIIAYSMGIIKIKGWAIIIWGNIN 215
Db      127        -           FDVWVWGAGYQYLLEALVRIIIIIKVHILLIGNIIIQGWKNLSVSVPI 173

Qy      216 YIPNIIIPIIIIVENYIIAIIEMIIFAFEV.IIIIIIIIANFIIVKIIIEVIIIIEIIIVII 275
Db      174 HIPQIS -        RYIIIIAQHISFVIIPIKIIISIIIIRVIIDFYVFIIIQFKAI   ANMHI 221

Qy      276 DSDIDSFSV 284
Db      222 DFIIDGHEL 230


RESULT 8
P96127
ID   P96127       PRELIMINARY;       PRT;    242 AA.
AC   P96127;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   Outer membrane protein.
GN   TROMP2
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID 160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN NICHOLS;
RX   MEDLINE 97175551; PubMed 9023206;
RA   Champion C.I., Blanco D.R., Exner M.M., Erdjument-Bromage H.,
RA   Hancock R.E., Tempst P., Miller J.N., Lovett M.A.;
RT   "Sequence analysis and recombinant expression of a 28-kDa Treponema
RT   Treponema pallidum subsp. pallidum rare outer membrane protein
RT   (Tromp2).";
RL   J. Bacteriol. 179:1230-1238(1997).
DR   EMBL; U65743; AAB47846.1; -.
SQ   SEQUENCE   242 AA;  27244 MW;  B6C71E31B6B62AF8 CRC64;

  Query Match             7.3%;  Score 125.5;  DB 2;  Length 242;
  Best Local Similarity   22.4%;  Pred. No. 0.19;
  Matches   55;  Conservative   41;  Mismatches  104;  Indels   55;  Gaps    8;

Qy       48 PSSIRLDLINYVLYV    YRQAIIVKIEDMVVLIINNW    SVLLIISARLQAYVKN 100
Db       21 PLAAQRFVNYQAYFIIIIGAS           IIQRII AWRAAGSRIIIIKGIIIAYFFI 70

Qy      101 SVVAIAVVEILSEFYAIIIIIIVEVIIIFYRQIAAMIM  --  IIKIIFYYIHIRIRQPL 155
Db       71 MPQAVRMAGSWQGKLEEAFFIVEIFKENRQGNNWLIIIPIKGGSDYEIIPLRGVVII       126

Qy      156 GKGLIINIKIMKRIKVIVIIIIIIIIIIIVIIIIMNIMIIAYSMGIIKIKGWAIIIWGNIN 215
Db      127    -   -   -   FDVWVWGAGYQYLLEALVRIIIIIKVHILLIGNIIIQGWKNLSVSVPI 173

Qy      216 YIPNIIIPIIIIVENYIIAIIEMIIFAFEV.IIIIIIIIANFIIVKIIIEVIIIIEIIIVII 275
Db      174 HIPQIS -        RYIIIIAQHISFVIIPIKIIISIIIIRVIIDFY        VFFDQFKALA 217

Qy      276 DSDID 280
Db      218 NMHID 222


RESULT 9
O83669
ID   O83669       PRELIMINARY;       PRT;    242 AA.
AC   O83669;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Outer membrane protein, putative.
GN   TP0663.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID 160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN NICHOLS;
RX   MEDLINE 98332770; PubMed 9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
DR   EMBL; AE001240; AAC65636.1; -.
DR   TIGR; TP0663; -.
KW   Complete proteome.
SQ   SEQUENCE   242 AA;  27257 MW;  B6C71E31B6B62AF8 CRC64;

  Query Match             7.1%;  Score 121.5;  DB 10;  Length 242;
  Best Local Similarity   22.0%;  Pred. No. 0.38;
  Matches   54;  Conservative   41;  Mismatches  105;  Indels   55;  Gaps    8;

Qy       48 PSSIRLDLINYVLYV    YRQAIIVKIEDMVVLIINNW     SVLLIPSARLQAYVKN 1
Db       21 LAAQRFVNYQAYFIIIIGAS           IIQRII AWRAAGSRIIIIKGIIIKYFFI 7

Qy      101 SVVAIAVVEILSEFYAIIIIIIVEVIIIFYRQIAAMIM      IIKIIFYYIHIRIRQPL 15
Db       71 MPQAVRMAGSWQGKLEEAFFIVEIFKENRQGNNWLIIIPIKGGSDYEIIPLRGVVSI      12

Qy      156 GKGLIINIKIMKRIKVIVIIIIIIIIIIIVIIIIMNIMIIAYSMGIIKIKGWAIIIWGNIN 21
Db      127    -   -   --FDVWVWGAGYQYLLEALVRIIIIIKVHILLIGNIIIQGWKNLSVSVPI 17

Qy      216 YIPNIIIPIIIIVENYIIAIIEMIIFAFEVIIIIIIIEVENIIIYVKIIEVIYGKISVSI 27
Db      174 HIPQIS --------RYIGSAQHISFVIIPIKIIISIIIIRVIIDFY        VFFDQFKALA 21

Qy      276 DSDID 280
Db      218 NMHID 222


RESULT 10
Q8U9V9
ID   Q8U9V9       PRELIMINARY;       PRT;    257 AA.
AC   Q8U9V9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Atu3616.
GN   ATU3616 OR AGR_L_2419.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID 176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 21608550; PubMed 11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,
```

[illegible]

```
DR   EMBL; [illegible]; AA[illegible].
DR   InterPro; IPR[illegible]; YcbB[illegible].
DR   InterPro; IPR[illegible]; Ch Mipept[illegible].
DR   Pfam; PF[illegible]; YcbB[illegible].
DR   SMART; SM[illegible].
SQ   SEQUENCE [illegible] AA; [illegible] MW; [illegible];

  Query Match [illegible];
  Best Local Similarity [illegible];
  Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Qy          4 KAKS[illegible]
Db         12 KVKN[illegible]
Qy         64 SGASGIVKPEDMVV[illegible]NNWSV[illegible]
Db         56 ---- - [illegible]LSSS[illegible]
Qy         [illegible] VKNSVVA[illegible]
Db        108 VDSQHSS[illegible]
Qy        [illegible]
Db        [illegible]
Qy        [illegible]
Db        [illegible]
Qy        [illegible]
Db        [illegible]
Qy        [illegible] EKISTAE[illegible]
Db        [illegible] EAI[illegible]SS[illegible]


RESULT [illegible]
Q9ZS[illegible]
ID   Q9ZS[illegible]   PRELIMINARY;   PRT;   [illegible] AA.
AC   Q9ZS[illegible];
DT   01 MAY 1999 (TrEMBLrel. [illegible], Created)
DT   01 MAY 1999 (TrEMBLrel. [illegible], Last sequence update)
DT   01 MAR 2002 (TrEMBLrel. [illegible], Last annotation update)
DE   Heat shock protein [illegible].
GN   HSP70.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; [illegible];
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=[illegible];
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GREEN SILIQUES;
RA   Kim [illegible] K.;
RL   Submitted (NOV-[illegible]) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=GREEN SILIQUES;
RA   Wunderlich K., [illegible]
RT   "Seed maturation-induced subset of heat shock protein mRNAs and heat
RT   shock element binding protein [illegible] dependent on ABI[illegible] in
RT   Arabidopsis thaliana.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN [illegible] FAMILY.
DR   EMBL; AJ0[illegible]; CAA[illegible].
DR   HSSP; P08107; 1DKX.
DR   InterPro; IPR[illegible]; Hsp[illegible].
DR   Pfam; PF[illegible]; HSP[illegible].
```

```
DR   PRINTS; PR00301; HEATSHOCK70.
DR   ProDom; PD000089; Hsp70; 1.
DR   PROSITE; PS00297; HSP70_1; 1.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   ATP binding.
SQ   SEQUENCE   650 AA;  71553 MW;  79A5B0[illegible] CRC64;

  Query Match             6.5%;  Score 111.5;  DB 10;  Length 650;
  Best Local Similarity   15.7%;  Pred. No. 7.6;
  Matches   6[illegible];  Conservative   58;  Mismatches  131;  Indels   9[illegible];  Gaps   14;

Qy         4[illegible] ELAI[illegible]NYV[illegible]NNW[illegible]AYVEN 100
Db         61 QVAMNIINTVFDAKRLIGRRYS  DPSVQADKSHWPFKVVSG[illegible]KK    104

Qy        1[illegible] SVVAFAVV  [illegible]AMIMPPEK  [illegible] 152
Db        1[illegible]  PMIVVNHEC[illegible]VLIKMR[illegible]SPVKNAVVIVPAYFNDSQR 159

Qy        1[illegible] [illegible]VLPE[illegible]YAY[illegible] 202
Db        1[illegible] QGIKDAIV[illegible]NVMRIINEPTAAAIAY[illegible]KNVLI[illegible]FDVSLLIIE 219

Qy        20[illegible]  FK[illegible]NYI[illegible]SKS 249
Db        22[illegible] [illegible]GNF[illegible] 279

Qy        2[illegible] HSSEVENE  [illegible] 28[illegible]
Db        28[illegible] LSSIAQTI[illegible]MEPVEK[illegible]AKM[illegible] 3[illegible]9

Qy        28[illegible] VYEIS[illegible]KI[illegible] 3[illegible]7
Db        3[illegible] VVVV[illegible]CK[illegible]EA 3[illegible]


RESULT 13
Q9Z102
ID   Q9Z102      PRELIMINARY;      PRT;  1865 AA.
AC   Q9Z102;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   RTX protein (APXIV).
GN   APXIVA OR APXIV VI.
OS   Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Actinobacillus.
OX   NCBI_TaxID=715;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN 4074;
RA   Kuhn R., Anderson I.[illegible], Segers R.A.P.M., MacInnes J.I., Nicolet J.,
RA   Frey J.;
RT   "Characterization of a new RTX determinant (apxIV) of Actinobacillus
RT   pleuropneumoniae.";
RL   Submitted (AUG 1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN 4074;
RA   Frey J., Segers R.P.;
RT   "Live attenuated bacteria of the species Actinobacillus
RT   pleuropneumoniae.";
RL   Patent number EP0875[illegible]
DR   EMBL; AF021919; AAD01698.1; -.
DR   EMBL; AX[illegible]
DR   InterPro; IPR001343; Hemlysn_Ca_bind.
DR   InterPro; IPR003355; RTXtoxin_N.
DR   Pfam; PF00353; hemolysinCabind; 11.
DR   PRINTS; PR00313; CABNDNGRPT.
DR   PROSITE; PS00330; HEMOLYSIN_CALCIUM; UNKNOWN_2.
SQ   SEQUENCE   1865 AA;  202129 MW;  A9076E[illegible]BDB16 CRC64;

  Query Match             6.4%;  Score 11[illegible];  DB 2;  Length 1865;
  Best Local Similarity   21.5%;  Pred. No. [illegible];
  Matches   68;  Conservative   51;  Mismatches  114;  Indels   84;  Gaps   1[illegible];

Qy         22 [illegible]GLAKGSKRAKPGELVLDFAL  [illegible]AKDPSSIE  [illegible]NYV[illegible]ASG[illegible]VKEE[illegible]M 7[illegible]
Db       1008 [illegible]ALPNG[illegible]FK[illegible]ENA[illegible]WK[illegible]AKA[illegible] 1[illegible]

Qy         7[illegible] VV[illegible]NNW[illegible]AKLQAYVEKGVVA  [illegible]EYA[illegible]VEVL[illegible] 1[illegible]
Db       1062    ---NW         AA[illegible]ARLFEA[illegible]VI[illegible]AESI[illegible]K[illegible]     1[illegible]

Qy        132 [illegible]SSAMIMPPEKIPPY[illegible] 1[illegible]
Db       1091 FKNNI[illegible]QKENNL[illegible]NE[illegible]KGHGQ[illegible]  IVYF[illegible] 11[illegible]

Qy        190 N[illegible]MEYAY[illegible]KEFKGWA[illegible]NYI[illegible]KMRFKAFEVSKS 24[illegible]
Db       1148 NN[illegible]NRAR[illegible]LKF      I[illegible]VNY AEVKFEKVENE[illegible]          M[illegible]YHE[illegible] 11[illegible]

Qy        250 [illegible]KVENE[illegible]VYEISE[illegible]ESLEELKAHEIE 30[illegible]
Db       1187 ISVIVKSFYSHVE--- YQFDKIEFA [illegible]RSIIRE[illegible]IKA[illegible]  [illegible]KIHAEW 12[illegible]

Qy        306 KRVLKLEEKISIAEGSE 322
Db       1240 ISILEGGKGNDILRGGY 1256


RESULT 14
Q9TL[illegible]4
ID   Q9TL[illegible]     PRELIMINARY;      PRT;  [illegible] AA.
AC   Q9TLJ4;
DT   01-MAY-2000 (TrEMBLrel. 1[illegible], Created)
DT   01-MAY-2000 (TrEMBLrel. 1[illegible], Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Inter-alpha-trypsin inhibitor heavy chain H[illegible].
GN   ITIH[illegible].
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   S[illegible]K A.;
RT   "The sequence of rabbit inter-alpha-trypsin inhibitor heavy chain H[illegible]
RT   cDNA.";
RL   Submitted (DEC 1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB035767; BAA88322.1; -.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00092; vwa; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
DR   PROSITE; PS50234; VWFA; 1.
SQ   SEQUENCE   9[illegible] AA;  10[illegible] MW;  [illegible] CRC64;

  Query Match             6.4%;  Score 1[illegible].5;  DB [illegible];  Length [illegible];
  Best Local Similarity   18.9%;  Pred. No. [illegible];
  Matches   78;  Conservative   67;  Mismatches  1[illegible];  Indels  137;  Gaps   1[illegible];

Qy          1 [illegible] [illegible]
Db        2[illegible] [illegible]ENM[illegible]INTV[illegible]  [illegible]FVE[illegible]   ----        3[illegible]4

Qy         6[illegible] YVYAGA[illegible]IVKEEE     MVV[illegible]NNWSVL[illegible]  AKL[illegible]A    YVEKGVVAIAVVKSE 1[illegible]
Db        31[illegible]        GD[illegible]KVL[illegible]GWKG[illegible]VFA[illegible]AAKIFV[illegible]A[illegible]AI     3[illegible]

Qy        112 SKKYA[illegible]VKVL       [illegible]SAMIMP[illegible]NKI 1[illegible]
Db        364    NLNGGLLRGIEILNNAQGNLFAVGKHAAIL      IMLIDGEP      IEGVI[illegible] 41[illegible]
```

Search completed: November
Job time : 97 secs

NFVEKIFSEKEESSPKN 341

SUMMARY

Description

flagellar filament
filamentous hemagg
hypothetical prote
N-acetylmuramoyl
probable ATP trans
xylulose kinase
hypothetical prote
dnaK type molecul
site-specific DNA
conserved hypothet
hypothetical prote
unknown protein

ALIGNMENTS

RESULT 1

flagellar filament outer layer protein (FlaA) [illegible]
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: [illegible]-Feb-1998 #sequence_revision [illegible]
C;Accession: [illegible]
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; [illegible]
son, D.; Peterson, J.; Kerlavage, A.R.; [illegible]
; Bowman, C.; Garland, S.; [illegible]
Nature 390, 580-586, 1997
A;Authors: Smith, H.; Venter, [illegible]
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi
A;Reference number: [illegible]
A;Accession: [illegible]
A;Status: preliminary [illegible]
A;Molecule type: DNA
A;Residues: 1-344 [illegible]
A;Cross-references: GB:AE[illegible]; NID:[illegible]; PIDN:AA[illegible]
A;Experimental source: strain B31
C;Superfamily: spirochete flagellar sheath protein

Query Match 99.0%; Score [illegible]; DB [illegible]; Length 344;
Best Local Similarity 98.8%; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

```
Qy    1 MKKKAKS[illegible]
Db    4 MKKKAKS[illegible]

Qy   61 YVYS[illegible]
Db   64 YVYSGASIVKPEIMVV[illegible]

Qy  121 [illegible]
Db  124 [illegible]

Qy  181 [illegible]
Db  184 [illegible]

Qy  241 FKAFFVSK[illegible]
Db  244 FKAFFVSKSHSSKLNFIIYVK[illegible]

Qy  301 AHEIEKKVIKI[illegible]NFVEKIF[illegible]KN 341
Db  304 AHEIEKKVIKI[illegible]NFVEKIFSEKEESSPKN 344
```

RESULT 2

```
flagellar filament outer layer protein flaA 1 precursor - syphilis spirochete
N;Alternate names: 37K endoflagellar sheath protein FlaA; minor endoflagellar protein; p
C;Species: Treponema pallidum subsp. pallidum [illegible]
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-2000
C;Accession: E71348; A40160; A37187; B27559; A32351
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; [illegible] Clayton, R.A.; [illegible] Chidambaram, M.; [illegible]
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: E71348
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-350 <CUL>
A;Cross-references: GB:AE001206; GB:AE000520; NID:g3322514; PIDN:AAC65235.1; PID:g3322516
A;Experimental source: strain Nichols
R;Isaacs, R.D.; Radolf, J.D.
Infect. Immun. 58, 2025-2034, 1990
A;Title: Expression in Escherichia coli of the 37 kilodalton endoflagellar sheath protei
A;Reference number: A40160; MUID:90307197; PMID:2194955
A;Accession: A40160
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-29 <ISA1>
A;Cross-references: GB:M63142; GB:M27915; NID:g155058
R;Isaacs, R.D.; Hanke, J.H.; Guzman-Verduzco, L.M.; Newport, G.; Agabian, N.; Norgard, M
Infect. Immun. 57, 3403-3411, 1989
A;Title: Molecular cloning and DNA sequence analysis of the 37 kilodalton endoflagellar
A;Reference number: A37187; MUID:90035499; PMID:2680972
A;Accession: A37187
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 21-350 <ISA2>
A;Cross-references: GB:M26525; NID:g155056
R;Blanco, D.R.; Champion, C.I.; Miller, J.N.; Lovett, M.A.
Infect. Immun. 56, 168-175, 1988
A;Title: Antigenic and structural characterization of Treponema pallidum (Nichols strain
A;Reference number: A91753; MUID:88085476; PMID:2447018
A;Accession: B27559
A;Molecule type: protein
A;Residues: [illegible]
R;Norris, S.J.; Charon, N.W.; Cook, R.G.; Fuentes, M.D.; Limberger, R.J.
J. Bacteriol. 170, 4072-4082, 1988
A;Title: Antigenic relatedness and N-terminal sequence homology define two classes of pe
A;Reference number: A32982; MUID:[illegible]; PMID:[illegible]
A;Accession: A32351
A;Molecule type: protein
A;Residues: 21-48,'R',50-51 <NOR>
C;Genetics:
A;Gene: TP0249
C;Superfamily: spirochete flagellar sheath protein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-350/Product: flagellar filament outer layer protein flaA 1 #status predicted <MA

  Query Match             21.4%;  Score 401.5;  DB 2;  Length 350;
  Best Local Similarity   28.5%;  Pred. No. 8.6e-19;
  Matches  106;  Conservative   66;  Mismatches  120;  Indels   65;  Gaps   11;

Qy          6 KSIL-FFLLSIVLFA[illegible]LAE[illegible]KEAL[illegible]VLIFAKLAE[illegible]INVVLYVY 63
Db         22 ESVLIDFAKLNAD[illegible]MADK[illegible]MIHNKK      IVLDYA[illegible]LAD      ISYID     63

Qy         64 SQA[illegible]VFFE[illegible]MVV[illegible]NNW[illegible]VLL[illegible]AYVEN[illegible]VVAFAVVK[illegible]YAGDIIGV 123
Db         64 ---[illegible]EQKALMKS[illegible]AVA[illegible]WEVVLNS[illegible]NEVAHAAG[illegible]VIIAFV[illegible]AGLKVIGV 117

Qy        124 RVLEPSY-SQSSAMIMPPFKIPEYS----QPSGN---        -QPL-CKQLIEN 162
Db        118 EVLEFEW[illegible]NAMIKFAEVIIAYLNVMA[illegible]QGNVQA[illegible]GVVKN 177

Qy        163 [illegible]AY[illegible] 222
Db        178 V[illegible]LK[illegible]AVNI[illegible]VVRE[illegible]EVRR[illegible]WRE[illegible]VK[illegible]

Qy        [illegible]
Db        238 KLVK[illegible]VVIA[illegible]RVVPE[illegible]AABA[illegible]VK[illegible]KAVIGIVRPEADE [illegible]

Qy        2[illegible] [illegible]KAHE[illegible]NEVEKIESEK [illegible]
Db        297 IKW       [illegible]QARREAE            RKEVEVAE[illegible]VIEV[illegible]QEK [illegible]


RESULT 3
A43824
periplasmic flagellar sheath protein precursor - Treponema hyodysenteriae
C;Species: Treponema hyodysenteriae
C;Date: 30-Jan-1993 #sequence_revision 30-Jan-1993 #text_change [illegible]
C;Accession: A43824
R;Koopman, M.B.; de Leeuw, O.S.; van der Zeijst, B.M.; Kusters, J.G.
Infect. Immun. 60, 2920-2925, 1992
A;Title: Cloning and DNA sequence analysis of a Serpulina (Treponema) hyodysenteriae
A;Reference number: A43824; MUID:92307926; PMID:[illegible]
A;Accession: A43824
A;Molecule type: DNA
A;Residues: 1-320 <KOO>
A;Cross-references: EMBL:X63[illegible]; NID:g4[illegible]; PIDN:CAA44[illegible]; PID:g[illegible]4
C;Comment: The flagella of this spirochete are reported to be composed of [illegible] wi
C;Genetics:
A;Gene: flaA
C;Superfamily: spirochete flagellar sheath protein
C;Keywords: flagellum; periplasmic space
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-320/Product: periplasmic flagellar sheath protein #status experimental <MAT>

  Query Match             19.3%;  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   28.8%;  Pred. No. [illegible];
  Matches   [illegible];  Conservative   [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy          9 LFFLLSIVLFA[illegible]AE[illegible]SKEAEPGEIVLI[illegible]INYVLYVYS[illegible]AS [illegible]
Db          4 LFVVLISIIIAASAY-[illegible]LINSI          [illegible]A             [illegible]NAENI[illegible]AGE [illegible]

Qy         [illegible]
Db         44 INEVVPVAENL[illegible]NWVVWLNEEAEI[illegible]VVINV[illegible]AKA[illegible]VEVHE [illegible]

Qy        12[illegible] [illegible]LEVE [illegible]
Db        104 LAAWNGTAL[illegible]VYFELEMY[illegible]YIE[illegible]VIHNV[illegible]WVY[illegible]VLYEVNI [illegible]

Qy        187 EDMNRMEYAY[illegible]KEE[illegible]WANEN[illegible]FKAEV [illegible]
Db        164 QNEF[illegible]EKSY[illegible]MGIVIIN[illegible]VRWENE[illegible]NVI[illegible]VKEDS[illegible] 222

Qy        247 [illegible]EKEKAHE[illegible]FK [illegible]
Db        22[illegible] YRIKE[illegible]YVVE[illegible]VVVV[illegible]FAIR[illegible]  [illegible]NEEK [illegible]NIES [illegible]

Qy        307 RVLKLREKISIAE 319
Db        277    AEIREQAELRL 287


RESULT 4
A32814
flagellar filament surface antigen - Spirochaeta aurantia
C;Species: Spirochaeta aurantia
C;Date: 28-Oct-1989 #sequence_revision 28-Oct-1989 #text_change [illegible]
C;Accession: A32814
R;Brahamsha, B.; Greenberg, E.P.
J. Bacteriol. 171, 1692-1697, 1989
A;Title: Cloning and sequence analysis of flaA, a gene encoding a Spirochaeta aurantia
A;Reference number: A32814; MUID:[illegible]; PMID:[illegible]
A;Accession: A32814
```

Qy 1 MKKKAKS
Db 1 M
Qy 195 MF YA
Db 241 MQVYA

RESULT
I41142
filamentous hemagglutinin - Haemophilus ducreyi
C;Species: Haemophilus ducreyi
C;Accession: I41142
R;Ward, C.K.; Lumbley, S.R.; Latimer
J. Bacteriol. 180
A;Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein
A;Accession: I41142
A;Status: preliminary
A;Molecule type: DNA
A;Cross-references: EMBL
C;Genetics:
A;Gene: fhaA

Query Match
Best Local Similarity
Matches 84; Conservative 44; Mismatches

Qy 4 KAKSIL
Db 12 KVKNCLVPVAENIKSACNGRGGNSKIALS
Qy 64 SGASSIVKPEEMVVSGINNWSV
Db 56

```
Qy     98 YKNSVVAPAVVKSESKRYA[illegible]GVRV[illegible]PSYSQSSAMIMPPFKIP-   FYSG--ESG 151
Db    198 VDSQHSSTKIYEQKINENSK[illegible]NSEKFFNIPK[illegible]SRTEST 167
Qy    152 NQFLGKGLIDNIKIMKEIKVSVYSLGYEIDLEVLFEDMNGMEYAYSM[illegible]WA[illegible] 210
Db    168 SQLVGK-IHANIQLQK[illegible]NIQGALEVA[illegible]EAF[illegible] 2[illegible]
Qy    211 WSNPNYIPNISSKEIKDDVENTLLA[illegible]SKMEFKAFKV-[illegible]--HSSKVKN[illegible]FYVK[illegible]KV 266
Db    214 IVNPNGIILNGVKIINI[illegible]KEVVS[illegible]HP--ENGLLSVKN[illegible] 256
Qy    267 LYDKLSVS-------IDSDID[illegible]ESVFKVYEISGIESLRKLKAHETF-----KKVLKLR 312
Db    257 IIDKGGVAINGLSHEEVVARNIDQKGKIIVAKTENQKSVN--PANI[illegible]AAGSLNINLKIR 314
Qy    313 EKISIAEGS 321
Db    315 EAIFISSGI 323
```

RESULT 8
140492
gntR type transcription regulator homolog yexb - Bacillus subtilis
N:Alternate names: hypothetical protein 8 (srfA operon)
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: 140492; F69766; S34990
R;Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van Sind
Mol. Microbiol. 8, 821-831, 1993
A;Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
A;Reference number: 140485; MUID:93360813; PMID:8355609
A;Accession: 140492
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-444 <RES>
A;Cross-references: EMBL:X70356; NID:g396480; PIDN:CAA49821.1; PID:g396486
A;Experimental source: ATCC 21332
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
A;Authors: Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
tech, J.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
Koetter, P.; Koningstein, G.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel
A;Authors: Lauber, J.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Y, M.; Ogawa, K.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlan
Rieger, M.; Rivolta, C.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
A;Authors: Schleich, S.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
akeuchi, M.; Tamakoshi, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
I.; Winters, P.; Wipat, A.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69766
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-444 <KUN>
A;Cross-references: GB:Z99105; GB:Z99106; GB:AL009126; NID:g2632653; PID:g2632657; NID:g
A;Experimental source: strain 168
C;Genetics:
A;Gene: yexb
C;Superfamily: hypothetical protein b1439

```
  Query Match             6.3%;  Score 109;  DB 1;  Length 444;
  Best Local Similarity   21.6%;  Pred. No. 2.5;
  Matches   75;  Conservative   42;  Mismatches  126;  Indels  104;  Gaps   15;

Qy     29 SKRAEPGELVLDFAELARDPSSIRLDLTNYVD----------------YVYSGASGI--- 69
Db     73 SKSGQPGP--IDFATSAPDP----DVFPYLDFQHCINKAIDTYKNDLFTYGTPKGLPSL 125
Qy     70 -------VKPEDMVVDLGINNWSVLL-            IPSAFLAYVKN [illegible]
Db    126 IRVLRKLLAIQQVPADERHIFITSGVQQALSLLCAMPFPNGKEKIAIEQPSYHLMVEQLF 18[illegible]
Qy    101 SVVAPAVVKSESKRYACDIIGVRVLPPSYSQSSAMIMPPFKIPFYSGESGNQFLGKGLI 1[illegible]
Db    186 TIGIPAIGVKRIEE-GLDIAKVEPLEGIFSIKFFYIMPRFHNP              LEGSLS 2[illegible]
Qy    1[illegible] DNIKIME[illegible]KVSVYSLGYEIDLEVLFEDMNGMEYAYSMGILKFKGWALLIWSNPNYIPNI 2[illegible]
Db    2[illegible] L[illegible]ARKAI              LLTAYDI 2[illegible]
Qy    2[illegible]
Db    2[illegible]
Qy    221 SSFIIKGFVENI[illegible]KMKEFKAFKVGK[illegible]          STAEGSI 2[illegible]
Db    275 SSHVI---YIKSFSKMMFPGLRVGAVLPEALIDIIA  [illegible]
Qy    281 [illegible]LSVFKVYEISGIL[illegible]EKLKAHETFK KVLKLEKI
Db    322 SQAALEIYLKSGMYGRHKEKIEDSYKERSLRLHQAIPIHRGLSGRI 3[illegible]
```

RESULT 9
F64385
hypothetical protein MJ0686 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-[illegible]
C;Accession: F64385
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannas
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: F64385
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-580 <BUL>
A;Cross-references: GB:U67515; GB:L77117; NID:g1499[illegible]; PIDN:AAB98681.1; PID:g149[illegible]
C;Genetics:
A;Map position: REV615409-613667

```
  Query Match             6.3%;  Score 1[illegible];  DB [illegible];  Length 5[illegible];
  Best Local Similarity   21.7%;  Pred. No. 4.4;
  Matches   56;  Conservative   38;  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy    112 SKRYACDIIGVRVLPPSTSQSSAMIMPPFKIPFYSGESGNQFLGKGLIDNIKIMKEIKV 171
Db    376 S[illegible]VSIST          -ESKMMKEIKEVSQWIENNNKI-IFVVENLKI  FIDY 41[illegible]
Qy    172 SVYSLGYEIDLEVLFEDMNGMEYAYSMGILKFKGWALLIWSNPN   YIPNISSKEIKD 228
Db    41[illegible] QAYSL[illegible]          KKVFSSGSSKYKVFNILKNKPMEIYKP            IN 4[illegible]
Qy    229 VENTLLASSKMEFKAF[illegible]          KVKSHSSKVKNEFYVK 2[illegible]
Db    4[illegible] [illegible]                              YVK [illegible]
Qy    2[illegible] [illegible]ISSKKSRK
Db    454 IAYTLEDHHIFPKGFLENKG[illegible]SESVFKVYEISGIESLRKLKAHETFKKVLKL 3[illegible]
Qy    2[illegible] DKVLYDKLSVSIDSDID---                                [illegible]
Db    508 EMIEIQKNKGLSEDFAVNKVKEILKGHINELMELIRNI[illegible]SKEIEENENKIEL [illegible]
Qy    312 REKISIAEGSFQNFVEKI 329
Db    567 REKL-------LEKI 575
```

RESULT 10
G64427
H+-transporting two-sector ATPase (EC 3.6.1.34) chain [illegible] - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 0[illegible]
C;Accession: G64427

[illegible] R.; Sutton, G.G.; Blake, [illegible] K.M.; Hurst, M.A.

[illegible] Smith, H.O.; Woese, [illegible] Methanococcus jannaschii [illegible]

[illegible] not shown

[illegible] N;AAB[illegible]

[illegible]

[illegible]

Result 12
A41322
N;[illegible]
N;Alternate names: major autolysin [illegible]
C;Species: Bacillus subtilis
C;Date: [illegible]
C;Accession: A41322; [illegible]
R;Kuroda, A.; Sekiguchi, [illegible]
J. Bacteriol. 173, 7304-7312, [illegible]
A;Title: Molecular cloning and sequencing [illegible]
A;Reference number: A41322; [illegible]
A;Accession: A41322
A;Molecule type: DNA
A;Residues: [illegible]
A;Cross-references: [illegible]
R;Lazarevic, V.; Margot, P.; [illegible]; Karamata, [illegible]
J. Gen. Microbiol. [illegible]
A;Title: Sequencing and analysis [illegible]
A;Reference number: A47[illegible]
A;Accession: [illegible]
A;Molecule type: DNA
A;Residues: [illegible]
A;Cross-references: [illegible]
A;Experimental source: strain [illegible]
A;Note: sequence extracted from NCBI [illegible]
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; [illegible]; Azevedo, V.; [illegible]
[illegible]; Bron, S.; Brouillet, S.; Bruschi, [illegible]; Caldwell, B.; [illegible]; Carter, N.M.; [illegible]
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; [illegible]
Nature 390, 249-256, 1997
A;Authors: [illegible]
[illegible]; Harwood, C.R.; Henaut, A.; Hilbert, H.; [illegible]
Koetter, P.; Koningstein, [illegible]; Krogh, S.; [illegible]
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; [illegible]
[illegible]; Park, S.H.; [illegible]
[illegible]
A;Authors: Schleich, S.; Schroeter, R.; [illegible]
[illegible]
[illegible]; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; [illegible]
A;Authors: [illegible]
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: [illegible]
A;Accession: [illegible]
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-[illegible] <KUN>
A;Cross-references: [illegible]
A;Experimental source: strain [illegible]
C;Comment: This protein enhances the activity of the N-acetylmuramoyl-L-alanine amidase
[illegible]

```
A;Gene: lytB
A;Start codon: TTG
C;Keywords: cell wall
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-705/Product: N-acetylmuramoyl-L-alanine amidase modifier #status predicted <MAT>

  Query Match             6.1%;  Score 105.5;  DB 2;  Length 705;
  Best Local Similarity   25.1%;  Pred. No. 8.7;
  Matches   76;  Conservative   82;  Mismatches  120;  Indels   75;  Gaps   11;

Qy         45 KEPSSIRLDLI NYVIYVVYGAS [illegible] VKPEDMVVI [illegible] NNWGVLLI [illegible] AKLQAYVKNSVV 103
Db        252 KIPGSIRYE [illegible] IK [illegible] KVVM [illegible] LI [illegible] AGKKN [illegible] 302

Qy        104 APAVVKG [illegible] K  KYA [illegible] VKVL [illegible] AMIM [illegible] NQF 154
Db        303 FVE [illegible] AAAK [illegible] KA [illegible] A [illegible] VEN [illegible] 355

Qy        155 I  GF [illegible] NIF [illegible] MF [illegible] FV [illegible] MNGM [illegible] KI [illegible] WA 207
Db        356 LKINS [illegible] KLN [illegible] NIKIY [illegible] GK [illegible] IVNE [illegible] 402

Qy        208 DLIWSNFNY [illegible] NISSK [illegible] YVEN YI [illegible] AFEV [illegible] VKN [illegible] V 261
Db        403 [illegible] NLM [illegible] AQTVAA [illegible] KI 454

Qy        262 K [illegible] KVLYDK [illegible] SV [illegible] IFK [illegible] SIAE [illegible] 321
Db        455 GIIVP [illegible] TA [illegible] WN [illegible] NKAV [illegible] TK [illegible] AA [illegible] 500

Qy        322 FQN 324
Db        501 SSN 503


RESULT 13
D96955
probable ATP transporter permease component CAC0454 [imported] - Clostridium acetobutyli
C;Species: Clostridium acetobutylicum
C;Date: 14 Sep 2001 #sequence_revision 14 Sep 2001 #text_change 14 Sep 2001
C;Accession: D96955
R;Noelling, J.; Breton, G.; Omelchenko, M.V.; Makarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D96955
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-832 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78 [illegible]; GSPDB:GN [illegible]
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0454

  Query Match             6.1%;  Score 105;  DB 2;  Length 832;
  Best Local Similarity   19.9%;  Pred. No. 12;
  Matches   66;  Conservative   63;  Mismatches  122;  Indels   80;  Gaps   [illegible];

Qy         39 LDFAHLARG [illegible] NNW [illegible] AKL [illegible] ATV 98
Db        41? L [illegible] PENIERNE [illegible] AVV [illegible] VA  V [illegible] VVI [illegible] VFNS [illegible] MK [illegible] 459

Qy         99 KN [illegible] VVAPAVVK [illegible] KPYA [illegible] VEVL [illegible] AMIMP [illegible] KIP 144
Db        460 K [illegible] N [illegible] FNE [illegible] ANKKLEH 518

Qy        145 FY [illegible] VLE [illegible] MN [illegible] YAY [illegible] MGI 200
Db        519 K [illegible] YFWP [illegible] A [illegible] YFA [illegible] Y  [illegible] NE [illegible] NEVK  [illegible] YVIKGKIDI [illegible] VLAFEK 567

Qy        201 LKFKGWA [illegible] IWSN [illegible] NY [illegible] VENY [illegible] KMEFKA [illegible] EVKN 25?
```

```
Db        568 IKYFGV [illegible] NNKVY [illegible] KN [illegible] 62?

Qy        257 E FYVKDLRVL  -  YDKLSVS [illegible] SVEKVY [illegible] SGIFS [illegible] KK [illegible]
Db        62? [illegible] KANK [illegible] VSKK [illegible] K [illegible]

Qy        301 AHLIEKVLE [illegible] SIAE [illegible] NEVEKIES 33?
Db        67? [illegible] VNYKYI [illegible] ESQYKQY  KIEF 701


RESULT 14
B70126
surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13 Feb 1998 #sequence_revision 13 Feb 1998 #text_change 1? Jan 200?
C;Accession: B70126
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
[illegible]
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:94 [illegible]
A;Accession: B70126
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1119 <KLE>
A;Cross-references: [illegible] NID:g2688 [illegible] PIDN:AA [illegible]
A;Experimental source: strain B31
C;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat
F;742-774/Domain: tetratricopeptide repeat homology #status atypical <TI1>
F;775-808/Domain: tetratricopeptide repeat homology <TI2>
F;809-842/Domain: tetratricopeptide repeat homology <TI3>
F;843-876/Domain: tetratricopeptide repeat homology <TI4>
F;911-943/Domain: tetratricopeptide repeat homology <TI5>
F;944-977/Domain: tetratricopeptide repeat homology <TI6>
F;978-1010/Domain: tetratricopeptide repeat homology #status atypical <TI7>
F;1011-1044/Domain: tetratricopeptide repeat homology <TI8>

  Query Match             6.1%;  Score 104.5;  DB 2;  Length 1119;
  Best Local Similarity   21.7%;  Pred. No. 20;
  Matches   58;  Conservative   44;  Mismatches  1[illegible];  Indels   [illegible];  Gaps   [illegible];

Qy         98 VKNGVVAPAVVKG [illegible] KKYAG [illegible] VKVL [illegible] SY [illegible] AMIM [illegible] 14?
Db         16 VENSVII  ENYEKAY [illegible]  VAK [illegible] KY [illegible] AM [illegible] ANSSPFFSK [illegible] 66

Qy        147 [illegible] EGL [illegible] KI [illegible]  MF [illegible] MN [illegible] M 192
Db         67 [illegible] SAN [illegible] KIK [illegible] NIK [illegible] NVNENF [illegible] ANNEY [illegible] NY [illegible] VGIEDEDIS 155

Qy        193 EYAYSMG [illegible] KFKGWADLIWSNFNY [illegible] NISSK [illegible] NY [illegible] ASSKM [illegible] KAFRV [illegible] SS 25?
Db        156 EFKKSKIPEKIK  -  - [illegible] PNINFKE [illegible] VNP [illegible] KN [illegible] 20?

Qy        253 KVENF [illegible] FYVKDLRVLYDKLSV [illegible] SVEKVY [illegible] IFS [illegible] KKI [illegible]
Db        207 K [illegible] NTN [illegible] NNNN [illegible] NSQKI [illegible]

Qy        300  -  -KAHEIPKRVLK  --  [illegible] FEKI [illegible] A EGSEQNF 32?
Db        267 SELSPF [illegible] KIYKPY [illegible] KKDLYL [illegible] KNF [illegible] NKFQKV 32?

Qy        326 VEF [illegible] SEKF [illegible] 338
Db        327 NE [illegible] ENSKNKEAS 33?


RESULT 15
T31105
hypothetical protein 2 - Haemophilus ducreyi
C;Species: Haemophilus ducreyi
```

[illegible]

```
[illegible] VKNSVVAFAVV 1[illegible]
[illegible] IKNFVKEM[illegible] 10[illegible]2
[illegible] 141
[illegible] NKTIV[illegible] 11[illegible]
[illegible] KVSVYS[illegible]YE 17[illegible]
[illegible] NVKNK[illegible] KVLLYTKIFE 11[illegible]
[illegible] 21[illegible]
[illegible] NNNEKKY[illegible] SAFEATEA 1[illegible]
[illegible] ASSKMFE 24[illegible]
[illegible] NIKVLK[illegible]KW 1[illegible]
[illegible] SESVFKVY 2[illegible]
[illegible] FKKKYSK 1[illegible]
[illegible] KSVIK[illegible] SF[illegible]NFV [illegible]
[illegible] KKVENV[illegible] VSFENLN 1[illegible]
```

[illegible] Seconds
[illegible] updates/sec

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
26	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
31	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
35	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
36	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
38	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
39	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
41	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
45	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

ALIGNMENTS

RESULT 1
LOCUS [illegible]
DEFINITION [illegible]
ACCESSION [illegible]
VERSION [illegible]
KEYWORDS
SOURCE [illegible]
ORGANISM [illegible]
REFERENCE [illegible]
AUTHORS [illegible]

```
TITLE       An unexpected flaA homolog is present and expressed in Borrelia
            burgdorferi
JOURNAL     J. Bacteriol. 179 (2), 552-556 (1997)
MEDLINE     97144545
 PUBMED     8990312
REFERENCE   2  (bases 1 to 1655)
AUTHORS     Ge,Y.
TITLE       Direct Submission
JOURNAL     Submitted ([illegible]) [illegible] West Virginia [illegible],
            Microbiology, HSCN, Morgantown, WV 26506, USA
FEATURES             Location/Qualifiers
     source          1..1655
                     /organism "Borrelia burgdorferi"
                     /strain "212"
                     /db_xref "[illegible]"
                     /clone "pW1 and pW2"
     CDS             <1..395
                     /function "unknown"
                     /note "[illegible]"
                     /codon_start 3
                     /transl_table 11
                     /protein_id "AAC44769.1"
                     /db_xref "GI:[illegible]"
                     /translation "[illegible]EFVAN[illegible]NLDLLVKDFLLG
                     SN[illegible]SFF[illegible]
                     IKIM[illegible]NYP[illegible]LVE[illegible]LKEQS"
     gene            473..1498
                     /gene "flaA"
     CDS             473..1498
                     /gene "flaA"
                     /function "putative flagellar filament outsheath protein"
                     /codon_start 1
                     /transl_table 11
                     /product "FlaA protein"
                     /protein_id "AA[illegible]"
                     /db_xref "GI:1579447"
                     /translation "M[illegible]VLFA[illegible]
                     ARDPSSTRLDLINYVDYVYSCASGIVKF[illegible]DMVVD[illegible]INNWSVLLTPGARLQAYVKNSV
                     VAPAVVE[illegible]SSAMIMF[illegible]
                     DN[illegible]GMF[illegible]
                     NI[illegible]SH[illegible]SFVFNF[illegible]
                     [illegible]RVL[illegible]QNFVEF[illegible]KP
                     EESSPKN"
     gene            1566..1655
                     /gene "cheA"
     CDS             1566..>1655
                     /gene "cheA"
                     /function "chemotaxis histidine kinase"
                     /codon_start 1
                     /transl_table 11
                     /product "CheA protein"
                     /protein_id "AAC44771.1"
                     /db_xref "GI:[illegible]"
                     /translation "MEI[illegible]VFFEEAQNLVD[illegible]"
BASE COUNT      560 a    182 c    308 g    605 t
ORIGIN

  Query Match             98.4%;  Score 1637;  DB 1;  Length 1655;
  Best Local Similarity   99.5%;  Pred. No. 4.5e-221;
  Matches 1655;  Conservative    0;  Mismatches    0;  Indels    8;  Gaps    1;

Qy          1 ATGATAATCTTTTTCAAAAAAGGTTTTTTATTTCATCTAGCAAGGGATTTGTTGCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATAATCTTTTTCAAAAAAGGTTTTTTATTTCATCTAGCAAGGGATTTGTTGCTA 60

Qy         61 ATTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAG 120

Qy        121 ATTTTCTTTTAGGAAGCAATGAGGGGTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTTCTTTTAGGAAGCAATGAGGGGTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGAT 180

Qy        181 [illegible] 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTATATTCTTTTTTAAAGAATGGAGTTTATTATGTAAAT[illegible]AAGAGAATTTTATG 240

Qy        241 ATTCTTTTAATAATGGTGATTATAATGAATCTTTTGATGTTAAG[illegible]AATCTTTTTGCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCTTTTAATAATGGTGATTATAATGAATCTTTTGATGTTAAG[illegible]AATCTTTTTGCTA 300

Qy        301 [illegible] 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 [illegible]TTTAATAAAAA[illegible]AAAAAAATTATTATT[illegible] 360

Qy        361 [illegible] 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 [illegible] 420

Qy        421 AATTAAAAAACGAAAATTTTATAAAAGATTTATATATAA[illegible]TATATGAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AATTAAAAAACGAAAATTTTATAAAAGATTTATATATAA[illegible]TATATGAAAAG 480

Qy        481 GAAAGCTAAAAGTATTTTATTTTTTTATTATC[illegible]AAGA[illegible]A 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAAGCTAAAAGTAT[illegible]A 540

Qy        541 [illegible] 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGATTAGCAGAGGGTTCTAAAAGGGTAGAGCTGGAGAATTAGTTTAGATTTTGGCTA 600

Qy        601 GCTTGCAAGAGATGCAAGTTCAACTAGA[illegible]A 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 [illegible]A 660

Qy        661 [illegible]A 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 [illegible]A 720

Qy        721 [illegible] 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 [illegible] 780

Qy        781 [illegible]AAAAA[illegible]TATTTTGGGGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 [illegible]GCTGTTGTTAAGA[illegible]AAAAAG[illegible]ATA[illegible]ATTTTGGGGGT 840

Qy        841 AAGA[illegible]CAAGT[illegible]AAT[illegible]TTAAAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGA[illegible]CAAG[illegible]AAT[illegible]TTAAAAT 900

Qy        901 [illegible]TTATT[illegible]AGGGAAA[illegible]AAT[illegible]AAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 [illegible]TTATT[illegible]AGGGAAA[illegible]AAT[illegible]AAT 960

Qy        961 [illegible]AAAA[illegible]ATGAAAGAAATTAA[illegible]TTGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TAAAACCATGAAAGAAATTAAGGTATCTGTTTATAGTTTAGGTATGAGATAGATTTGA 1020

Qy       1021 GGTTTTATTTGAAGATATGAATGGCATGGAATA[illegible]TTAAAGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GGTTTTATTTGAAGATATGAATGGCATGGAATATG[illegible]TTAAAGTT 1080

Qy       1081 [illegible]AAAGG[illegible]GGCTGATTTAAT[illegible]AAAT[illegible] 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 [illegible]AAAGGTGGGCTGATTTAATTGGTCAAATG[illegible]AAATATAGCA[illegible] 1140

Qy       1141 AATTATTAAAGACGATGTTCCAAATTATCC[illegible]AAGTAAAATGAGATTTAAGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATTATTAAAGACGATGTTCCAAATTATCC[illegible]AAGTAAAATGAGATTTAAGGC 1200

Qy       1201 TTTTAGAGTTTCAAAGTCACACAGTTCAAAAGTTAAAAATTTTATCTTTTATGTTAAAGA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTAGAGTTTCAAAGTCACACAGTTCAAAAGTTAAAAATTTTATCTTTTATGTTAAAGA 1260

Qy       1261 TTTAAGAGTTCTTTATGATAACTAAGTGTTT[illegible]GATATTGACAGTGAGTC 1320
```

```
                     GNFVTTLBAKSVEL[illegible]NKG[illegible]
                     V"
     gene            5706..6146
                     /gene="BB0672"
     CDS             5706..6146
                     /gene="BB0672"
                     /note="similar to [illegible] percent identity; [illegible]
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="chemotaxis response regulator (cheY-3)"
                     /protein_id="AAC67021.1"
                     /db_xref="GI:2688604"
                     /translation="M[illegible]
                     FTSEGEN[illegible]
                     VIM[illegible]RAKV[illegible]VK"
     gene            complement(6187..6702)
                     /gene="BB0673"
     CDS             complement(6187..6702)
                     /gene="BB0673"
                     /note="[illegible] percent identity; [illegible]
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="[illegible] hypothetical protein"
                     /protein_id="AAC67020.1"
                     /db_xref="GI:2688603"
                     /translation="M[illegible]NAW[illegible]EEFAL[illegible]
                     EEF[illegible]
                     FLF[illegible]LEIY[illegible]
                     ISNEMI[illegible]"
     gene            complement(6699..7745)
                     /gene="BB0674"
     CDS             complement(6699..7745)
                     /gene="BB0674"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BB0674"
                     /protein_id="AAC67027.1"
                     /db_xref="GI:2688610"
                     /translation="MVF[illegible]
                     LFDI[illegible]
                     IILNPIMLSLF[illegible]IYYNSKEENL[illegible]
                     IISSSF[illegible]INEKAYIKEKYP[illegible]
                     [illegible]
                     [illegible]
                     EFVEF[illegible]IK"
     gene            complement(7739..8602)
                     /gene="BB0675"
     CDS             complement(7739..8602)
                     /gene="BB0675"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BB0675"
                     /protein_id="AAC67026.1"
                     /db_xref="GI:2688609"
                     /translation="MNIFTYFFYLYB[illegible]
                     TFIFNFFFSFIKF[illegible]YVKFKT[illegible]VLSCIEA[illegible]
                     FNFVKVNSSLRKEFFYNFNI[illegible]INKLFDGKNSYTIYLIQKENKALYSSD[illegible]
                     KNYIKIL[illegible]ESTSKAISNDTD[illegible]
                     YFT[illegible]VKG[illegible]
                     ELSEIAKVNKW"
     gene            complement(8599..9263)
                     /gene="BB0676"
     CDS             complement(8599..9263)
                     /gene="BB0676"
                     /note="similar to [illegible]
                     PID:1222502 PID:1204815 percent identity: 25.58;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoglycolate phosphatase (gph)"
                     /protein_id="AAC67019.1"
                     /db_xref="GI:2688602"
                     /translation="M[illegible]
                     LV[illegible]
                     MNKLN[illegible]NAIMILFL
                     [illegible]
                     IK"
     gene            9504..11114

  Query Match             72.8%;  Score 1216.[illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   98.[illegible]%;  Pred. No. [illegible]e-[illegible];
  Matches 1246;  Conservative    [illegible];  Mismatches    [illegible];  Indels   [illegible];  Gaps    [illegible];

Qy        461 TAAATT[illegible]ACTAA[illegible]AAAAATAA[illegible]AAAAAA[illegible]ATATATA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAAATTTTACTAATAAAAATAATTAAAAAACGAAAATTTTATAAAAGATTTATATATATA 60

Qy        [illegible] [illegible] 518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGGAGTTGGTTTACATGAAAAGGAAAGCTAAAAGTATTTTATTTTTTTATTATCTACTG 120

Qy        [illegible] [illegible]AAGAGA[illegible]GATTAGCA[illegible] [illegible]
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 [illegible]AAGAGACTGATGGATTAGCAG[illegible]TAAAA[illegible] 180

Qy        [illegible] AATTAGT[illegible]AGATT[illegible]AAGAGA[illegible] 638
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATTAGTTTTAGATTTTGGGGAGTTGGAAGAGAT[illegible]AATTTCAATTAGATTTGATTTTA 240

Qy        639 CAAATTATGTTGATTATGTATATTGGG[illegible]GCTT[illegible]AAT[illegible]AAGA[illegible] 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAATTATGTTGATTATGTATATTGGG[illegible] 300

Qy        699 TTGTAGATCTTGGGATAAATAATTGGAGCGTTTTACTTACT[illegible]TTCT[illegible]AAGGTTGCAGG 758
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGTAGATCTTGGGATAAATAATTGGAGCGTTTTACTTAC[illegible]AAGGTTGCAGG 360

Qy        75[illegible] [illegible]TTAC[illegible]TAAAAATT[illegible]ACT[illegible]AAAAAGGTACG 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 [illegible]TTAC[illegible]TAAAAAT[illegible]CAGT[illegible]AAAAAGGTACG 420

Qy        819 CAGGTGATACTATTTTGGGGGTAAGAGTTTTGTTTCCAAGCTATTCTCAATCATCTGCTA 878
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        421 CAGG[illegible]GATACTAT[illegible]TAGGGGTAAGAGTTTTG[illegible]AAG[illegible]AATCAT[illegible]A 480

Qy        87[illegible] [illegible]GATTATG[illegible]ATTTAAAA[illegible] 9[illegible]8
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGATTATGCCACCATTTAAAATTCCTTTTATTCAGGGGAAAGTGGCAATCAATTTTTAG 54[illegible]

Qy        9[illegible] [illegible]AAAGG[illegible]TATT[illegible]AATAT[illegible]AAAA[illegible]CAT[illegible]AAAGAAA[illegible] [illegible]
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 [illegible]AAAGG[illegible]TATTGATAA[illegible]ATTAAAA[illegible]CATGAAAGAAA[illegible]AGT[illegible] [illegible]

Qy        99[illegible] TAGGGTATGAGATAGAT[illegible]TTGAGG[illegible]TTATTTGAAGATATGAAT[illegible]ATGGAATATGCTT 10[illegible]8
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAGGGTATGAGATAGATCTTGAGG[illegible]TTTATTTGAAGATATGAATGGCATGGAATATGCTT 660

Qy       1059 AT[illegible]TAT[illegible]TTTAAA[illegible]TTAAAG[illegible]TTAA[illegible]AAAT[illegible]AACT 1118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ATTCTATGGGTACTTTAAAGTTTAAAGGGTGGGCTGATTTAATTTGGTCAAATGCTAACT 720

Qy       11[illegible] ATAT[illegible]TAATATATCA[illegible]AGAATTATTAAAGA[illegible]AAA[illegible]ATCCTCTTGCTT 11[illegible]8
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATATTCCTAATATATCATCCAGAATTATTAAAGACTAT[illegible]AAATTATCCTCTTGCTT 780

Qy       1179 CAAGTAAAATGAGATTTAAGG[illegible]TTTAGAGT[illegible]AAA[illegible]AAAAT[illegible]AAAA 12[illegible]8
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAAGTAAAATCAGATTTAAGG[illegible]TTTACAGT[illegible]AAAATATAAA 84[illegible]

Qy       12[illegible] AT[illegible]TCAT[illegible]TTTTAT[illegible]TTAAAGATTTAAGATT[illegible]AATAA[illegible]AA[illegible]GATA[illegible] 1298
```

complete cds.

Borrelia

and Seppala,I.

burgdorferi sens.

Lyme

and Seppala,I.

Helsinki

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

```
Db    901 GCACAGAAA[illegible]AAAAGAGTTTTAAAGCTTAGAGAAAAAATTTCTATGCTGAAGG 960

Qy   1433 [illegible]TTTCCAAAACTTTTAGAAAAGATTGAGATGAAAAACTGAAGAATCATCTCGAAA 1492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TCTTTCCAAAACTTTTAGAAAAGATTGAGATGAAAAACTGAAGAATCATCTCGAAA 1020

Qy   1493 AATTAT 1498
Db   1021 AATTAT 1026


RESULT 4
AY057223
LOCUS       AY057223                1026 bp    DNA     linear   BCT 31-OCT-2001
DEFINITION  Borrelia afzelii A91 flagellin A (flaA) gene, complete cds.
ACCESSION   AY057223
VERSION     AY057223.1  GI:16555361
KEYWORDS    .
SOURCE      Borrelia afzelii.
  ORGANISM  Borrelia afzelii
            Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE   1  (bases 1 to 1026)
  AUTHORS   Panelius,J., Lahdenne,P., Saxen,H., Heikkila,T. and Seppala,I.
  TITLE     Recombinant flagellin A proteins from Borrelia burgdorferi sensu
            stricto, B. afzelii, and B. garinii in serodiagnosis of Lyme
            borreliosis
  JOURNAL   J. Clin. Microbiol. 39 (11), 4013-4019 (2001)
  MEDLINE   21528965
   PUBMED   11682523
REFERENCE   2  (bases 1 to 1026)
  AUTHORS   Panelius,J., Lahdenne,P., Saxen,H., Heikkila,T. and Seppala,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2001) Haartman Institute, University of Helsinki,
            Haartmaninkatu 3, Helsinki 00290, Finland
FEATURES             Location/Qualifiers
     source          1..1026
                     /organism="Borrelia afzelii"
                     /strain="A91"
                     /db_xref="taxon:29518"
     gene            1..1026
                     /gene="flaA"
     CDS             1..1026
                     /gene="flaA"
                     [illegible]
                     /transl_table=11
                     /product="[illegible]"
                     /protein_id="AAL23583.1"
                     /db_xref="GI:16555362"
                     /translation="[illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     [illegible]
                     EESSSEN"
BASE COUNT    329 a    [illegible]    247 g    [illegible] t
ORIGIN

  Query Match             [illegible];  Score [illegible];  DB 1;  Length 1026;
  Best Local Similarity   [illegible];  Pred. No. [illegible];
  Matches  [illegible];  Conservative    0;  Mismatches   75;  Indels   [illegible];  Gaps   [illegible];

Qy    473 ATGAAAAGGAAAGCTAAAAGTA[illegible]CACTG[illegible]AA 532
Db      1 ATGAAGAGGAAAGCGAAAAGTATTTTATTTTTTTATTATCCACTGCTCTTTTGCTCAA 60

Qy    533 GAGACTGAT[illegible]TGGAGAATTAGTCTTAGAT 592
Db     61 GAGACTGAT[illegible]TGGAGAATTGGTTTTAGAT 120

Qy    593 [illegible]
Db    121 [illegible]

Qy    653 [illegible]
Db    181 [illegible]

Qy    713 [illegible]
Db    241 [illegible]

Qy    773 [illegible]
Db    301 [illegible]

Qy    833 [illegible]
Db    361 [illegible]

Qy    893 [illegible]
Db    421 [illegible]

Qy    953 [illegible]
Db    481 [illegible]

Qy   1013 [illegible]
Db    541 [illegible]

Qy   1073 [illegible]
Db    601 [illegible]

Qy   1133 [illegible]
Db    661 [illegible]

Qy   1193 [illegible]
Db    721 [illegible]

Qy   1253 [illegible]
Db    781 [illegible]

Qy   1313 [illegible]
Db    841 [illegible]

Qy   1373 [illegible]
Db    901 [illegible]

Qy   1433 [illegible]
Db    961 [illegible]

Qy   1493 AATTAC 1498
Db   1021 AATTAC 1026


RESULT 5
AY057224
LOCUS       AY057224                1029 bp    DNA     linear   BCT 31-OCT-2001
DEFINITION  Borrelia garinii 40 flagellin A (flaA) gene, complete cds.
ACCESSION   AY057224
VERSION     AY057224.1  GI:16555363
KEYWORDS    .
SOURCE      Borrelia garinii.
  ORGANISM  Borrelia garinii
```

[illegible]

RESULT 6
AE001167
LOCUS AE001167 [illegible]
DEFINITION Borrelia burgdorferi [illegible] complete [illegible]
ACCESSION AE001167 AE [illegible]
VERSION AE001167 [illegible]
KEYWORDS .
SOURCE Borrelia burgdorferi.
ORGANISM Borrelia burgdorferi
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.
REFERENCE 1 (bases 1 to [illegible])
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., [illegible] White,O., Ketchum,K.A., [illegible] Gwinn,M., [illegible] Peterson,J., Kerlavage,A.R., Quackenbush,J., [illegible] Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., [illegible] Venter,J.C. et al.
TITLE Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi
JOURNAL Nature [illegible]
MEDLINE [illegible]
PUBMED [illegible]
REFERENCE 2 (bases 1 to [illegible])
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., [illegible] White,O., Ketchum,K.A., [illegible] Gwinn,M., Dougherty,B., [illegible] Fleischmann,R.D., Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., Gocayne,J., Weidman,J., [illegible] Watthey,L., McDonald,L.

Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K., Hatch,B., Smith,H. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted ([illegible]) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1..[illegible]
/organism="Borrelia burgdorferi"
/db_xref="taxon:139"
gene 84..1217
/gene="BB0656"
CDS 84..1217
/gene="BB0656"
/note="similar to [illegible] GB:[illegible] percent identity: 25.62; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oxygen independent coproporphyrinogen III oxidase, putative"
/protein_id="AAC67009.1"
/db_xref="GI:2688591"
/translation="[illegible]"
gene complement(1204..1890)
/gene="BB0657"
CDS complement(1204..1890)
/gene="BB0657"
/note="similar to GB:L77117 PID:1500500 percent identity: 39.53; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribose 5-phosphate isomerase (rpiA)"
/protein_id="AAC67008.1"
/db_xref="GI:[illegible]"
/translation="[illegible]"
gene 2024..2785
/gene="BB0658"
CDS 2024..2785
/gene="BB0658"
/note="similar to [illegible] GB:U00096 PID:[illegible] PID:1786970 percent identity: 59.76; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoglycerate mutase (gpmA)"
/protein_id="AAC67007.1"
/db_xref="GI:[illegible]"
/translation="[illegible]"
gene complement(2853..4418)
/gene="BB0659"
CDS complement(2853..4418)
/gene="BB0659"
/note="similar to GB:L77117 SP:Q57659 PID:1591243 percent identity: 33.14; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="lysyl tRNA synthetase"
/protein_id="AAC67006.1"
/db_xref="GI:[illegible]"
/translation="[illegible]"
gene complement(4506..5432)
/gene="BB0660"
CDS complement(4506..5432)
/gene="BB0660"
/note="similar to GB:[illegible] percent identity: 37.05; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="GTP-binding protein (era)"
/protein_id="AAC67005.1"
/db_xref="GI:[illegible]"
/translation="[illegible]"
gene 5467..5838
/gene="BB0661"
CDS 5467..5838
/gene="BB0661"
/note="hypothetical protein; identified by Glimmer; putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0661"
/protein_id="AAC67015.1"
/db_xref="GI:2688597"
/translation="[illegible]"
gene 5848..6258
/gene="BB0662"
CDS 5848..6258
/gene="BB0662"
/note="hypothetical protein; identified by Glimmer; putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0662"
/protein_id="AAC67014.1"
/db_xref="GI:2688596"
/translation="[illegible]"
gene 6278..6727
/gene="BB0663"
CDS 6278..6727
/gene="BB0663"
/note="hypothetical protein; identified by Glimmer; putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0663"
/protein_id="AAC67013.1"
/db_xref="GI:2688595"
/translation="[illegible]"
gene complement(6736..7419)
/gene="BB0664"

```
"MLRNIMLIA ... TALISNBKNIQYBKKIKK
... NSYKNSKIKEIIIIVKVFLIN
... VYIVKAFEFANGIILLAIDF
... LKIKINSIKESNNEIKEIQIE

"MEI ... FISFVILSKIKIFFEDLF

AAAAA ... IAITIIIIFSTA 609
... IA IAAAAA 966
... AAIAAIAI 1824
... AAIAAIIIAI 1794
... AAAAAAAA 1874
... IIAAIIIIAT 1994
AAT AAAAIIIIAIAAAAIAI A AIAIA 1616

ORGANISM  Borrelia burgdorferi.
          Bacteria; Spirochaetes; ... Borrelia
          Borrelia burgdorferi group.
REFERENCE 1  (bases 1 to 342)
  AUTHORS
  TITLE   A cheA cheW operon in Borrelia burgdorferi ...
          disease
  JOURNAL ...
  MEDLINE ...
  PUBMED  ...
REFERENCE 2  (bases 1 to 342)
  AUTHORS ...
  TITLE   Direct Submission
  JOURNAL Submitted ...
          Mycology, ... FRANCE
FEATURES           Location/Qualifiers
     source        1..342
                   /organism="Borrelia burgdorferi"
                   /strain="..."
                   /db_xref="taxon:..."
                   /clone_lib="..."
     gene          1..202
                   /gene="..."
     CDS           ..
                   /gene="..."
                   /codon_start=1
                   /transl_table=11
                   /protein_id="AAC..."
                   /db_xref="GI:..."
                   /db_xref="SPTREMBL:..."
                   /translation="..."
                   SIQNEVEKIISEKPTISSIEN"
     repeat_region 205..256
                   /note="hairpin"
     RBS           ...
     gene          ...
                   /gene="cheA"
     CDS           ...
                   /gene="cheA"
                   /function="chemotactic ..."
                   /codon_start=1
                   /transl_table=11
                   /product="histidine kinase"
                   /protein_id="AAC..."
                   /db_xref="GI:..."
                   /db_xref="SWISS-PROT:..."
                   /translation="MEI...N"
BASE COUNT ...
ORIGIN

  Query Match ...
  Best Local Similarity ...
  Matches ...

Qy       1297 AGAIICIAIAIIAAIAIIAII ... AAAIIIIAIAAIAIAIIIAAIIAAI
Db          1 AGAIICIAIAIIAAIAIIAII ... AAAIIIIAIAAIAIAIIIAAIIAAI

Qy       1357 ... AAAAAAI
Db         61 ... AAAAAAI

Qy       1417 ... AAAA ... AAAAA ... A 147
Db        121 ... AAAA ... AAAA ... A 1..

Qy       1477 AGAAICAIIIAAAAAIIAIIIIAAAIIAAIAIIAAAIIAIIAAAA ...
Db        181 AGAAICAIIIAAAAAIIAIIIIAAAIIAAIAIIAAAIIAIIAAAA ...

Qy       1537 IIEKKYSSNAIAIIAAAAIAAIA ... AAAA ...
Db        241 II        AIAIAIIAAAAIAAIA ... AAAA ...
```

```
Qy     1597 [illegible]AAAAT 1636
Db      293 [illegible]AAAAT 332


RESULT 8
BBU28962
LOCUS       BBU28962                2700 bp    DNA     linear   BCT 16-JAN-1998
DEFINITION  Borrelia burgdorferi histidine kinase (cheA) gene, complete cds.
ACCESSION   U28962
VERSION     U28962.1  GI:1113814
KEYWORDS    .
SOURCE      Borrelia burgdorferi.
  ORGANISM  Borrelia burgdorferi
            Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE   1  (bases 1 to 2700)
  AUTHORS   Trueba,G.A., Old,I.G., Saint Girons,I. and Johnson,R.C.
  TITLE     A cheA cheW operon in Borrelia burgdorferi, the agent of Lyme
            disease
  JOURNAL   Res. Microbiol. 148 (3), 191-200 (1997)
  MEDLINE   98430936
   PUBMED   9765799
REFERENCE   2  (bases 1 to 2700)
  AUTHORS   Trueba,G.A. and Johnson,R.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1995) Gabriel A. Trueba, Microbiology, University
            of Minnesota, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
COMMENT     On Dec 12, 1995 this sequence version replaced gi:981522.
FEATURES             Location/Qualifiers
     source          1..2700
                     /organism="Borrelia burgdorferi"
                     /strain="CI-1"
                     /db_xref="taxon:139"
     gene            95..2689
                     /gene="cheA"
     CDS             95..2689
                     /gene="cheA"
                     /function="involved in chemotaxis"
                     /codon_start=1
                     /transl_table=11
                     /product="histidine kinase"
                     /protein_id="AAB68655.1"
                     /db_xref="GI:1113815"
                     /translation="[illegible]
                     [illegible]
                     DEQKDIKE"
BASE COUNT      988 a    278 c    507 g    927 t
ORIGIN

  Query Match             5.5%;  Score 164.4;  DB 1;  Length 2700;
  Best Local Similarity   95.3%;  Pred. No. 2.9e-14;
  Matches  183;  Conservative    0;  Mismatches    1;  Indels    8;  Gaps    1;

Qy     1472 GCTGAAGAATCATCTGGAAAAATTAGGTTTAAATTAATATGTAAAGCTACCTAAAAGGT 1531
Db        1 GCTGAAGAATCATCTGGAAAAATTAGGTTTAAATTAATATGTAAAGCTACCTAAAAGGT 60

Qy     1532 TTGCTTTTERROYSSNACATATTAAAATAATAGGAAATAGTATATGGAAATATTAGATTTG 1591
Db       61 TTGCTTT        ACATATTAAAATAATAGGAAATAGTATATGGAAATATTAGATTTG 1..

Qy     1592 GAAAATGAAGAGATTTTAGGAG[illegible]GAAGAAG[illegible]AAAAT[illegible]AGATATC[illegible] 1651
Db      113 GAAAATGAAGAGATTTTAGGAGTTTTTTTGAAGAAGTCAAAATTTTAGATATCTT 1..

Qy     1652 GAAGAGAATATT 1663
Db      173 GAAGAGAATATT 184


RESULT 9
A24079
LOCUS       A24079                 1140 bp    mRNA    linear   PAT 0- MAR-199-
DEFINITION  T.hyodysenteriae 4- kDa sheath protein gene.
ACCESSION   A24079
VERSION     A24079.1  GI:904392
KEYWORDS    .
SOURCE      Brachyspira hyodysenteriae.
  ORGANISM  Brachyspira hyodysenteriae
            Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
REFERENCE   1  (bases 1 to 1140)
  AUTHORS   Koopman,M.B.H. and Kusters,J.G.
  TITLE     Treponema hyodysenteriae vaccine
  JOURNAL   Patent: EP 0534526-A 1 31-MAR-1993;
            DUPHAR INTERNATIONAL RESEARCH B.V
FEATURES             Location/Qualifiers
     source          1..1140
                     /organism="Brachyspira hyodysenteriae"
                     /strain="C5"
                     /db_xref="taxon:159"
                     /cell_line="CBS 513.91"
     35_signal       27..32
                     /note="putative"
     10_signal       50..55
                     /note="putative"
     RBS             89..91
                     /note="putative"
     CDS             103..1065
                     /codon_start=1
                     /transl_table=11
                     /product="endoflagellar sheath protein"
                     /protein_id="AAB01715.1"
                     /db_xref="GI:904393"
                     /db_xref="SWISS-PROT:P32520"
                     /translation="MKKIFVV[illegible]
                     [illegible]
                     AETEQAELK[illegible]AAN[illegible]AQ"
     sig_peptide     103..159
BASE COUNT      390 a    152 c    233 g    36- t
ORIGIN

  Query Match             6.5%;  Score 108.6;  DB 6;  Length 1140;
  Best Local Similarity   49.4%;  Pred. No. 2.5e-06;
  Matches  [illegible];  Conservative    0;  Mismatches  [illegible];  Indels    0;  Gaps    0;

Qy      581 [illegible]A 64.
Db      136 TTTATGCTGCATCT[illegible]AAAC[illegible]AACA 195

Qy      641 AATTATGT[illegible] 70.
Db      196 GGTAATGTGATAATTACAAC[illegible]GGAGAAGG[illegible]CA 255

Qy      701 GTAGATCTTGGGATAAATAATTGGAGCGTTTTACTTACT[illegible]CAGGCT 76.
Db      256 GAAAATCTTTATAAT[illegible]AAG[illegible] 31.

Qy      761 TACGTTAAAAATTCAGT[illegible]AAGAG[illegible]AAAAAGGTACGGA 820
```

[illegible]

THE NETHERLANDS

[illegible]

CDS [illegible]
sig_peptide [illegible]
mat_peptide [illegible]
repeat_region [illegible]
/rpt_type=INVERTED
repeat_region [illegible]
/rpt_type=INVERTED
BASE COUNT [illegible]
ORIGIN

Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible] Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

Qy 581 [illegible]
Db 136 [illegible]
Qy 641 [illegible]
Db 196 [illegible]
Qy 701 [illegible]
Db 256 [illegible]
Qy 761 [illegible]
Db 316 [illegible]
Qy 821 [illegible]
Db 376 [illegible]
Qy 878 [illegible]
Db 436 [illegible]
Qy 935 [illegible]
Db 496 [illegible]
Qy 995 [illegible]
Db 556 [illegible]
Qy 1055 [illegible]
Db 616 [illegible]
Qy 1115 [illegible]
Db 676 [illegible]
Qy 1175 [illegible]
Db 733 [illegible]

```
Qy     1235 AAAAA[illegible]A[illegible]AAA[illegible]AAGA[illegible]ATGATAA[illegible]AA[illegible]A 1264
Db      793 [illegible]GAGA[illegible]AT[illegible]AAA[illegible]AA[illegible]ATGA[illegible] 822

Qy     1295 ATAGAT[illegible]GATAT[illegible]GACAG[illegible]AG[illegible] 1321
Db      853 TTGAAGAAGATAT[illegible]GACGATGAAG[illegible] 879

RESULT 11
PFMAL1P3/c
LOCUS       PFMAL1P3              67970 bp    DNA     linear   INV 15-DEC-1999
DEFINITION  Plasmodium falciparum MAL1P3, complete sequence.
ACCESSION   AL031746
VERSION     AL031746.3  GI:6594243
KEYWORDS    HTG.
SOURCE      Plasmodium falciparum 3D7.
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 67970)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Dec 16, 1999 this sequence version replaced gi:[illegible].
            For more information about this sequence or the Malaria Project,
            see http://www.[illegible] [illegible]
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
FEATURES             Location/Qualifiers
     source          1..67970
                     /organism="Plasmodium falciparum 3D7"
                     /strain="3D7"
                     /db_xref="taxon:[illegible]"
                     /chromosome="1"
     gene            complement(join(1748..[illegible],2748..[illegible],[illegible]..[illegible]))
                     /gene="MAL1P3.01"
     CDS             complement(join(1748..[illegible],2748..2818,[illegible]..[illegible]))
                     /gene="MAL1P3.01"
                     /note="MAL1P3.01, conserved hypothetical protein, len: [illegible]
                     aa, similarity: UPF0006 family eg to
                     [illegible] YEA[illegible] ([illegible] aa), fasta
                     scores: opt: [illegible], E(): [illegible], ([illegible]% identity in 27[illegible] aa
                     overlap)"
                     /codon_start=1
                     /product="conserved hypothetical protein, UPF0006 family"
                     /protein_id="CAB63556.1"
                     /db_xref="GI:[illegible]"
                     /translation="[illegible]
                     [illegible]
                     IIKCDENTIIKKENEDYNIA"
     misc_feature    complement(2599..2610)
                     /gene="MAL1P3.01"
                     /note="potential splice acceptor sequence"
     misc_feature    complement(2742..2747)
                     /gene="MAL1P3.01"
                     /note="potential splice [illegible] sequence [illegible]"
     misc_feature    complement(2849..2861)
                     /gene="MAL1P3.01"
                     /note="potential splice acceptor sequence"
     misc_feature    complement(2984..2989)
                     /gene="MAL1P3.01"
                     /note="potential splice [illegible] sequence [illegible]"
     gene            5005..5496
                     /gene="MAL1P3.02"
     CDS             5005..5496
                     /gene="MAL1P3.02"
                     /note="MAL1P3.02, hypothetical protein, len: [illegible] aa
                     contains possible signal sequence"
                     /codon_start=1
                     /product="hypothetical protein MAL1P3.02"
                     /protein_id="CAB63557.1"
                     /db_xref="GI:6594245"
                     /translation="MKLLNN[illegible]
                     [illegible]
                     LNN"
     misc_feature    8020..10389
                     /note="possible cen1, region of very high (A+T) content"
     gene            14884..20352
                     /gene="MAL1P3.03"
     CDS             14884..20352
                     /gene="MAL1P3.03"
                     /note="MAL1P3.03, putative ABC transporter, len: [illegible] aa"
                     /codon_start=1
                     /product="putative ABC transporter"
                     /protein_id="CAB63558.1"
                     /db_xref="GI:6594246"
                     /translation="[illegible]
                     GDIQELPNICRNFLVPYYASKLL[illegible]
                     [illegible]
                     [illegible]SELSHLLFMK"
     rRNA            [illegible]..31[illegible]
                     /gene="rRNA"
     misc_RNA        2[illegible]896..31[illegible]
                     /gene="rRNA"
                     /note="region containing small subunit, 5.8S and large
                     subunit rRNA genes and spacer regions"
     gene            complement(join(31966..32476,3267[illegible]..[illegible]))
                     /gene="MAL1P3.04"
     CDS             complement(join(31966..[illegible],[illegible]..[illegible]))
                     /gene="MAL1P3.04"
                     /note="MAL1P3.04, conserved hypothetical membrane protein,
                     len: 203 aa. similarity: P. falciparum chromosome 2
                     PFB0110W, [illegible]96126 predicted integral membrane protein ([illegible]
                     aa), fasta scores: opt: [illegible], E(): 4.3e-15, ([illegible]% identity
```

[illegible]

"MSVL...YYNYIIESKIKENNNIKIIN ... FEHYENIIKINELSINS ... FYKNKE"

[illegible]

"MEIKMN ... NKEMISYEKILWYSEN ... ERMELIVLEMENNEN ... FEISIEFHMINNVEKIEII ... AAIYVEEN"

[illegible]

"MNVI ... NEIKIESNOLIKNONIILNKSEIS ... NGNLKNAHNNNE ... EFKKAKKEKVKKHVVKNVIEIE ... KSKVEEKNLSIQ ... KKEESKEVEEESKEV ... AFEEEDAFEEEIAFEE ... KKIKRNIKRNAKI"

[illegible]

```
Db   8406 ATIATAATTATATATTTAATATTTATTTAATTTATTAATATATATAATATTTTAT 8
Qy    ... ATTAATATA...AAT...ATAAAAA...AAAAATTTATAT...ATAAAAT
Db   8346 TTATTAATTATATTATTAATTTATATATTTAATATTTATTTAA TTATTTAATATAAT 8
Qy    ... ATTTTTTTTATTAATTAATTATTTTTTTT...AATTTATTTAATAT
Db   ....  ATT.ATT...AAT.AT.AT.AAT.AT...
Qy    ... TTTTATATT...AAATAATTTATTTTATTAT TAAATTTTTAATATAATTTTATT
Db   8226 ATTATTAATTTATATTTTTAATATTTTATTTTATTTTATTATT AATATAATATTTTA
Qy    ... ATTTTTTAATAATTTTATTATAATTAAT...AATTTTTTTA
Db   8166 TTTATTTAATTATATTATTATTTTTTTAATTTATATATTTAATATATTATTTTATTAT
Qy    ... TTTTTTTAATAAAAAAAA...AAAAAAAA...ATTATTATT
Db   8106 TTATTATTTAATTTAT...ATTTAATATATATATATATAATA AAAATATAA AATA
Qy    ... TT...AAA...AAAATTATTTTTATTAATAAAAAT
Db   8048 TTTTTTAATTATATTTATTAATATAAATTTTATTATAATTAATTATTTATTAATAAAATAAAT
Qy    ... AATTAAAAATTTAAATTTTTATAAAATATTTT...TTTTTATTTAAAT
Db   7988 ATATAATTATATATATAATTATTAAAATTTAATT...AATATATTTATTTTTATATATAATATT
Qy    481 GAAAGTTAAAATTATTTTATTTTTTTTATTATTTATTTTT
Db   7928 ATTAATATATATATTAAATTATTTATTATATATTATTTTTTT 7...
```

RESULT 12
AC...86/c
LOCUS AC... [illegible] DNA [illegible]
DEFINITION Homo sapiens BAC [illegible]
ACCESSION AC...
VERSION AC... GI:...1474
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 17...)
AUTHORS Guistan...,J.H. and Waterston,R.H.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 9906...
PUBMED 9847074
REFERENCE 2 (bases 1 to 17...)
AUTHORS [illegible] W...,M. and [illegible] A.
TITLE The sequence of Homo sapiens BAC [illegible]
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 17...)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (.. MAR ....) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 17...)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (.. MAY ....) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 17...)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (.. ... .) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```
REFERENCE   6  (bases 1 to 173957)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 173957)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 11, 2000 this sequence version replaced gi:4753246.
                  ---- Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu/gsc
               Contact: sapiens@watson.wustl.edu
                  ----- Summary Statistics
               Center project name: H_NH0155J05

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The position of this clone was established as part of a
            collaboration between the Human Chromosome Y Mapping Project
            (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,
            and David C. Page at the Whitehead Institute for Biomedical
            Research, Cambridge MA) and the Washington University Genome
            Sequencing Center, St. Louis MO.

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1 8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or from P. de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu).
            VECTOR: pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION.
            The clone sequenced to the right is RP11-160K17. Actual start of
            this clone is at base position 1 of RP11-155J5, actual end is at
            base position 173957 of RP11-155J5.
FEATURES             Location/Qualifiers
     source          1..173957
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Y"
                     /map="Y"
                     /clone="RP11-155J5"
                     /clone_lib="RPCI-11"
     repeat_region   976..1007
                     /rpt_family="AT_rich"
     repeat_region   1281..1435
                     /rpt_family="Alu"
     repeat_region   1438..1582
                     /rpt_family="(TA)n"
     repeat_region   1621..2154
                     /rpt_family="(TA)n"
     repeat_region   2160..7755
                     /rpt_family="L1"
     repeat_region   7758..10708
                     /rpt_family="L1"
     repeat_region   11258..11377
                     /rpt_family="L2"
     repeat_region   11334..12039
                     /rpt_family="(TATATG)n"
     repeat_region   12046..12950
                     /rpt_family="L1"
     repeat_region   15316..15479
                     /rpt_family="(TA)n"
     repeat_region   23465..23506
                     /rpt_family="AT_rich"
     repeat_region   23633..23653
                     /rpt_family="AT_rich"
     repeat_region   23961..23996
                     /rpt_family="(TG)n"
     repeat_region   24034..24234
                     /rpt_family="Alu"
     repeat_region   27018..27131
                     /rpt_family="Alu"
     repeat_region   27132..27154
                     /rpt_family="(CAA)n"
     repeat_region   27155..27312
                     /rpt_family="Alu"
     repeat_region   27324..27617
                     /rpt_family="Alu"
     repeat_region   29398..29681
                     /rpt_family="Alu"
     repeat_region   29682..29720
                     /rpt_family="(CA)n"
     repeat_region   29721..29736
                     /rpt_family="Alu"
     repeat_region   35566..35796
                     /rpt_family="Retroviral"
     repeat_region   35775..36379
                     /rpt_family="Retroviral"
     repeat_region   36749..36868
                     /rpt_family="MER4 group"
     repeat_region   37699..37719
                     /rpt_family="AT_rich"
     repeat_region   38833..38860
                     /rpt_family="AT_rich"
     repeat_region   39860..40054
                     /rpt_family="MIR"
     repeat_region   40141..40187
                     /rpt_family="(CA)n"
     repeat_region   40832..41426
                     /rpt_family="L2"
     repeat_region   41431..41523
                     /rpt_family="Alu"
     repeat_region   41555..41587
                     /rpt_family="T_rich"
     repeat_region   41589..43483
                     /rpt_family="L1"
     repeat_region   43510..43651
                     /rpt_family="Alu"
     repeat_region   43652..43952
                     /rpt_family="L2"
     repeat_region   44279..49197
                     /rpt_family="L1"
     repeat_region   49198..49332
                     /rpt_family="Alu"
     repeat_region   49333..50487
                     /rpt_family="L1"
     repeat_region   51792..52073
                     /rpt_family="Alu"
     repeat_region   52074..52105
                     /rpt_family="(CA)n"
     repeat_region   52192..52482
                     /rpt_family="MER1_type"
     repeat_region   52511..52829
                     /rpt_family="MER4-group?"
     repeat_region   53366..53426
```

[illegible]

```
REFERENCE   [illegible]
  AUTHORS   [illegible]
            and [illegible]
  TITLE     [illegible]
  JOURNAL   Submitted [illegible]
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     [illegible]
            For more information about this sequence [illegible]
            see http://www.sanger.ac.uk/[illegible]
            sequence is unfinished [illegible]
            [illegible]
            * NOTE: this is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          [illegible]
                     /organism="[illegible]"
                     /strain="[illegible]"
                     /[illegible]="[illegible]"
                     /chromosome="4"
BASE COUNT   [illegible]
ORIGIN

  Query Match             [illegible]
  Best Local Similarity   [illegible]
  Matches  [illegible]

[illegible]

RESULT [illegible]
```

```
MP015677
LOCUS       MP015677                1128 bp    DNA     linear   INV 09-FEB 1996
DEFINITION  Myrmecia pilosula H187 133 mitochondrion cytochrome b gene, partial
            cds.
ACCESSION   U15677
VERSION     U15677.1  GI:576756
KEYWORDS    .
SOURCE      Myrmecia pilosula.
  ORGANISM  Mitochondrion Myrmecia pilosula
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
            Formicidae; Myrmeciinae; Myrmecia.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Crozier,R.H., Dobric,N., Imai,H.T., Graur,D., Cornuet,J.M. and
            Taylor,R.W.
  TITLE     Mitochondrial-DNA sequence evidence on the phylogeny of Australian
            jack-jumper ants of the Myrmecia pilosula complex
  JOURNAL   Mol. Phylogenet. Evol. 4 (1), 20-30 (1995)
  MEDLINE   95346030
   PUBMED   7620633
REFERENCE   2  (bases 1 to 1128)
  AUTHORS   Crozier,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1994) Ross H. Crozier, School of Genetics, La
            Trobe University, Bundoora, Victoria 3083, Australia
FEATURES             Location/Qualifiers
     source          1..1128
                     /organism="Myrmecia pilosula"
                     /organelle="mitochondrion"
                     /strain="H187 133"
                     /db_xref="taxon:14619"
                     /tissue_type="head and thorax"
                     /dev_stage="adult worker"
     CDS             <2..733
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome b"
                     /protein_id="AAA66971.1"
                     /db_xref="GI:576757"
                     /translation="TWGAIVITNLVSTIFLGSDIIQWLWGGYGVSNATLNRFYSLHF
                     ILPFMILIMVIMHLYFLHGTGSSNPLNINSNLNKIIFNIYYSTKDMLGFIILLFITL
                     LIILMTPYLFSDPDNFIPANPLVTPIHIQPEWYFLFAYSILRSIPNKLGGVIALFSSI
                     ILYFLPMLISKMNSSSFYPLSQFLIWLFFMIFMILIWLGSQVIEHPYIIIAQFFSMLY
                     FSNFILLKLLYKMWLYLIYDKIINS"
     misc_feature    734..1128
                     /note="intergenic space"
BASE COUNT      415 a    138 c     63 g    529 t
ORIGIN

  Query Match              5.1%;  Score 84.8;  DB 3;  Length 1128;
  Best Local Similarity   50.4%;  Pred. No. 0.0054;
  Matches  233;  Conservative    0;  Mismatches  227;  Indels    2;  Gaps    1;

Qy         50 ATTTTTTTTAATTTAAGATATTTAAGAGATGAATAAAATTTGAAAGATAATTTAGATCT 109
Db        650 ATATTTTTTAATTTTATTTTATTAAAATTTTATATAAAATATGAAATTATTT  TATT 709

Qy        110 TTTAGTAAAAGATTTTCTTTTAGGAAGCAATGAGGGGTTTTCTTTGGGTTTTTATTAAG 169
Db        710 TATGATAAATTTTTTAATAATTAAAAATAATTTTACTTAAATATTTTATAAATTAAAAAT 769

Qy        170 TGATTCAAGATTTTTATATTCTTTTTTAAAGAATGGAGTTTATTATGTAAATCTTTCAAC 229
Db        770 TTTTATAAATTTAAATAAATAATTTTATAAACTTAAATTTATTTATTATTTATTTTATTT 829

Qy        230 AGAATTTTATGATTCTTTTAATAATGGGATTATAATGAATCTTTGATGTTAAGGTCAA 289
Db        830 TTTTTTAAATATTATTTTTTAAATATTATTTTTAAATATTATTTTAAATATTATTTTTAA 889

Qy        290 TCTTTTTGCTATGTCTTTAATAAAAACAATGCGCTTTAACTATCCGGTAAGATAAAAAA 349
Db        890 ATATTATTTTTAAATATTATTTTTAAATATTATTTTTAAATATTATTTTTAAATATTATT 949

Qy        350 CATTATTATTCTTGTTCAAGGGTTATTCTAAATGATCAAATTTATAAATTAAATTTTA 409
Db        950 TTTAAATATTATTTTTAAATATTATTATTTATTTTTATAATTTTATATATTAAATATA 1009

Qy        410 TAATAAAAATAATTAAAAAATAAAATTTTATAAAAGATTTATATATAAAGATTTTTT 469
Db       1010 TAAATAAGATATATTTATAATAAATAAATTTATTATTATTATATAAAAAATTAAATTATA 1069

Qy        470 TACATGAAAAGGAAAGTAAAAGTATTTTATTTTTTTATTA 511
Db       1070 TAAATTTTATTTAAAATTTTAAATTTTAATAAATAATTTTA 1111


RESULT 15
AC096750/c
LOCUS       AC096750             159618 bp    DNA     linear   HTG 01-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-426F15 from 4, complete sequence.
ACCESSION   AC096750 AC068637
VERSION     AC096750.3  GI:16418260
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159618)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 159618)
  AUTHORS   
  TITLE     The sequence of Homo sapiens BAC clone RP11-426F15
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 159618)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 159618)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 159618)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 25, 2001 this sequence version replaced gi:16259141.
            ------------ Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu/gsc
               Contact: sapiens@watson.wustl.edu
            ------------ Summary Statistics
               Center project name: H_NH0426F15
               Drafting Center: WIBR
            ------------

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
```

```
Qy        4 ATAATTTTTTTCAAAAAATTTTTTTATTTTCATTTCAGAAGGGATTTGTTGCTAATT 63
Db    11568 ATATTATATATATAAAATATATATATTATATATATAATATGTATATATTTTATAAAATAT 11509

Qy       64 TAAGATAT----TTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAA 119
Db    11508 ATATATTTTATATATATATATTATAAAATATATATATGAAATATATATATTTTATATATA 11449

Qy      120 GATTTTCTTTTAGGAAGCAATGAGGGGTTTTCTTTTGGGTTTTTAT      TAAGTGATTC 175
Db    11448 AAATATATATTATGTAAAATATATATATTTATGTAAAATATATATAATATATATTATAC 11389

Qy      176 AAGATTTTTATATTTTTTTTTAAAGAATTGAGTTTATTATGTAAATCTTTCAAGAGAATT 235
Db    11388 ATAATATATATAATATATATTATAAAAATATATATATATATAAAATATATATAATATAT 11329

Qy      236 TTATTATTTTTTTAATAATGGTGATTATAATGAATTTTTTGATGTTAAGGTCAATCTTTT 295
Db    11328 ATTATATAAAATATATGTAATATATAAAATATATATATTATATAATATATATATTATATA 11269

Qy      296 TGCTATGTCTTTAATAAAAACAATGCGCTTTAACTATCCTGGTAAGATAAAAAAGATTAT 355
Db    11268 TAACATATATATTATATAAAATATATATATTATATAAAATATATATTATATATAATATAT 11209

Qy      356 TATTCTTGTTGAAGGGTGTATCTTAAAGGAGCAAAGTTGATAAATTAAATTTTACTAATA 415
Db    11208 ATTATATAAAATATATATTATATATAAAATATATATTATATAAAATATATATTATATAAA 11149

Qy      416 AAAATAATTAAAAAACGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTACATG 475
Db    11148 ATATATATTATATATAAAATATATTATATATAAAATATATATTATATAAAATATATATTA 11089

Qy      476 AAAAGGAAAGTTAAAAGTATTTTATTTTTTTATTATTCACTGTTCTTTTTGTTCAAGAG 535
Db    11088 TATATAAAATATATATATTATATAAAATATATAATATTTAATATACATATTATATAAAAT 11029

Qy      536 ACTGATGGATTAGTAGAGGGTTCTAAAAGGGTAGAGTTGGAGAATTAGTCTTAGATTTT 595
Db    11028 GTATAATATGTAATATACATATTATATAAAATATATAATATGTAAT       ATACATATT 10974

Qy      596 GCCGAGCTTGCAAGAGATCCAAGTTCAACTAGACTTGATCTTACAAATTATGTTGATTAT 655
Db    10973 ATATAAAATATATAATATAAAATATATAATATATAATATATAATATAAAATATATAATAT 10914

Qy      656 CTATATTCGGGCGCTTCTGGTATTGTTAAGCCGGAAGATAT-GGTTGTAGATCTTGGGAT 714
Db    10913 ATAATATAAAATATATAAAATATATATAATATATAAAATATATAAAATATATAATATATA 10854

Qy      715 AAATAATTGGAGCGTTTTACTTACTCCTTCTGCAAGTTTGCAGGCTTACTTTAAAAATTC 774
Db    10853 AAATATATAAAATATATAATATATATTATATAAAATATATAATATATATATAAAATATAT 10794

Qy      775 AGTTGTTGCGCCGCTTGTTGTTAAGAGTGAGTCAAAAAGGTACGCAGGTGATACTATTTT 834
Db    10793 AATATATATATATAAAATATATAATACATATTATATAAAATATATAATATATAAATGTATA 10734

Qy      835 CGGGGTAAGAGTT          ---TTGTTTCCAAGCTATTCTCAATCATCTGCTATGATTATG 886
Db    10733 ATACATATTATATAAAATATATAATATATAATATATAATATAAAATATATATAATACATA 10674

Qy      887 CCACCATTTAAAATTCCTTTTTATTCAGGGGAAAGTGGCAATCAATTTTTAGGCAAAGGT 946
Db    10673 TTATATATTATATAAAATATATATAATATATATAAAATATATGTAATATATAATATATATAA 10614

Qy      947 CTTATTGATAACATTAAAACTATGAAAGAAATTAAGGTATCTGTTTATAGTTTAGGGTAT 1006
Db    10613 AATATATATTATACATAATATATAATATATATAAAATATTTATATATAATATATATTAC 10554

Qy     1007 GAGATAGATCTTGAGTTTTATTTGAAGATATGAATGGATGGAATATGCTTATTCTATG 1066
Db    10553 ATTATATATAAAATATATATATTATATATTATATATGTTATATAATATATATATTTTATA 10494

Qy     1067 GGTACTTTAAAGTTTAAAGG[illegible]ATTTAA[illegible]AAAT[illegible]AACTATAT[illegible] 11[illegible]
Db    10493 TAATATATTATATATAATATA[illegible]TATATAAT[illegible]TATATATTTA 1[illegible]

Qy     1127 AATATATCATCAGAATTATTAAAGACGATGTTTCAAATTAT[illegible]TTTCAATTAAA 1186
Db    10433 TATAATATATTATATATAATATATATATTTTATATAATATATATATATATTATATTTTAT 10374

Qy     1187 ATCAGATTTAAT[illegible]TATATTTTAAATTAT[illegible] 1246
Db    10373 ATAATATATTATATATAATATATATATTTTATATATTATATTATATATAATATATATATT 10314

Qy     1247 TTTTATGTTAAAGATTTAAGATTTTTTTATGATAATTTAAGTTTTTAATAGATTTTGAT 1306
Db    10313 TATATATTATATATATTATATTTAATATATATTTATATAAAATATTTATATATAATATA 10254

Qy     1307 ATTCAGAGTGAGTTTTATTTAAAGTTATGA[illegible]AAA 1366
Db    10253 TTTATATATAAAATATATTATATATAATATATTTTATA[illegible]AAAATA 10194

Qy     1367 TTAAAG[illegible]AAAA[illegible]AAA[illegible]AAAAAATTTTTATTTTT 1426
Db    10193 TTATATATTATATATAAAATATATAGAAAATATATTATATATTATATATAAAATATATAT 10134

Qy     1427 GAAGGTTTTTTCAAAACTTTGTAGAAAAGATTGAGATGAAAAATTTGAAGAATTATTT 1486
Db    10133 AAAATATATTATATATTGTATATAAAATATATATAAAATACATTATATATTGTATATAAA 10074

Qy     1487 CCGAAAAATTAGGTTTAAATTAATATGTAAAGCTACTTAAAAGGTTTGCTTTTRRYSSN 1546
Db    10073 ATATATATAAAATATATTATATATATGATATAATATATAAAATATATAATATATAAAAAT 10014

Qy     1547 ACATATTAAAATAATATGAAATA[illegible]AAA[illegible] 1606
Db    10013 ATATATTATATGTAAAATATAAAATATATAAAATA[illegible] 9954

Qy     1607 TTAGGAGTTTTTTTTGAATAATTTAAAATTTTTTATATATTTTTGAATATAATAT 1662
Db     9953 ATATAATATATATA[illegible]
```

Search completed: November 4, 2002, 22:15:33
Job time : 5143.24 secs

SUMMARIES

Description

FLAA [illegible] borrelia bu[illegible]
FLAA [illegible] spirochaeta
FLAA [illegible] treponema p
FLAA [illegible] P32520 treponema h
[illegible] Q08792 bacillus su
[illegible] Q58099 methanococ
VAT[illegible] MET[illegible] methanococ
[illegible] bacillus su
[illegible] sus scrofa
[illegible] bacillus s
[illegible] lycopersico
[illegible] mesocricetu
[illegible] Q45882 clostridium
[illegible] Q58943 methanococ
[illegible] trypanosoma
[illegible] plasmodium
[illegible] helicobacte
[illegible] mus muscul
[illegible] mycoplasma
[illegible] lycopersico
[illegible] saccharomyc
[illegible] P14450 torpedo cal
[illegible] saccharomyc
[illegible] pisum sati
[illegible] arabidopsis
[illegible] saccharomyc
[illegible] saccharomyc
[illegible] saccharomyc
[illegible] staphylococ
[illegible] borrelia bu
MSP[illegible] PLAFF P13819 plasmodium
MSP[illegible] PLAFM [illegible] plasmodium
[illegible] escherichia

34	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
35	[illegible]	[illegible]	[illegible]	1	RENI HUMAN	[illegible] homo sapien
36	[illegible]	[illegible]	[illegible]	1	AMY[illegible]	[illegible]
37	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
38	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
39	[illegible]	[illegible]	[illegible]	1	[illegible] HUMAN	[illegible]
40	[illegible]	[illegible]	[illegible]	1	[illegible]	[illegible]
41	[illegible]	[illegible]	[illegible]	1	MERM [illegible]	[illegible]
42	[illegible]	[illegible]	[illegible]	1	M[illegible]	[illegible]
43	[illegible]	[illegible]	[illegible]	1	K[illegible] HUMAN	[illegible] homo sapien
44	[illegible]	[illegible]	[illegible]	1	SYN[illegible]	[illegible]
45	[illegible]	[illegible]	[illegible]	1	ASA[illegible]	[illegible]

ALIGNMENTS

RESULT 1
FLAA BORBU
ID FLAA BORBU STANDARD; PRT; [illegible] AA.
AC [illegible]
DT [illegible] MAY [illegible] (Rel. [illegible], Created)
DT [illegible] MAY [illegible] (Rel. [illegible], Last sequence update)
DT [illegible] (Rel. [illegible], Last annotation update)
DE Flagellar filament outer layer protein precursor (Sheath protein).
GN FLAA OR BB[illegible].
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=[illegible];
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=212;
RX MEDLINE=97144[illegible]; PubMed=[illegible];
RA Ge Y., Charon N.;
RT "An unexpected flaA homolog is present and expressed in Borrelia
RT burgdorferi.";
RL J. Bacteriol. 179:552-556(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=[illegible] / B31;
RX MEDLINE=980[illegible]; PubMed=[illegible];
RA Fraser C.M., Casjens S., Huang W.M., [illegible], Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., [illegible], Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
RN [3]
RP SEQUENCE OF [illegible]-[illegible] FROM N.A.
RC STRAIN=212;
RX MEDLINE=9843[illegible]; PubMed=[illegible];
RA [illegible]old I.G., Truelta [illegible];
RT "A [illegible] Borrelia burgdorferi [illegible]
RT disease.";
RL Res. Microbiol. 148:191-20[illegible](1997).
CC -!- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC -!- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF
CC REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS SEVERAL
CC ANTIGENICALLY RELATED POLYPEPTIDES.
CC -!- SUBCELLULAR LOCATION: Periplasmic flagellum.
CC [illegible]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```
DR   EMBL; [illegible]; AA[illegible].1; -.
DR   EMBL; AE[illegible]168; AA[illegible]25.1; ALT INIT.
DR   EMBL; X[illegible]; CAA6[illegible].1; -.
DR   TIGR; BB0668; -.
KW   Flagella; Periplasmic; Signal; Complete proteome.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    341       FLAGELLAR FILAMENT OUTER LAYER PROTEIN.
FT   CONFLICT    254    255       EQ -> VK (IN REF. 1).
FT   CONFLICT    317    318       MP -> IA (IN REF. 1 AND 3).
SQ   SEQUENCE   341 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             99.0%;  Score 1701;  DB 1;  Length 341;
  Best Local Similarity   98.8%;  Pred. No. 1.2e-117;
  Matches  337;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKKKAKS[illegible]AELARKPSSTRLDLINYVD[illegible] 60
Db          1 MK[illegible]AELAR[illegible] 60

Qy         61 [illegible]VVA[illegible]AVV[illegible]YA[illegible] 120
Db         61 [illegible]VVA[illegible]AVV[illegible]YA[illegible] 120

Qy        121 DGVRVLFP[illegible]AMIM[illegible]NIKIM[illegible] 180
Db        121 DGVRVLFPSY[illegible]AMIM[illegible]NIKIMEE[illegible] 180

Qy        181 DLEVLFEDMN[illegible]WA[illegible]LIW[illegible]SR[illegible]ME 240
Db        181 DLEVLFEDMN[illegible]WA[illegible]LIW[illegible]SR[illegible]ME 240

Qy        241 [illegible]ESV[illegible] 300
Db        241 [illegible]LSV[illegible] 300

Qy        301 AHEIFFKV[illegible]EN 341
Db        301 AHEIFFKVLE[illegible]EN 341


RESULT 2
FLAA_SPIAU
ID   FLAA_SPIAU     STANDARD;      PRT;   337 AA.
AC   P21982;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Flagellar filament outer layer protein precursor (Sheath protein).
GN   FLAA.
OS   Spirochaeta aurantia.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
OX   NCBI_TaxID=147;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89155489; PubMed=2921247;
RA   Brahamsha B., Greenberg E.P.;
RT   "Cloning and sequence analysis of flaA, a gene encoding a Spirochaeta
RT   aurantia flagellar filament surface antigen.";
RL   J. Bacteriol. 171:1692-1697(1989).
RN   [2]
RP   REVISIONS.
RA   Greenberg E.P.;
RL   [illegible] data to [illegible]
RN   [3]
RP   SEQUENCE OF 22-[illegible].
RC   STRAIN=M1;
RX   MEDLINE=91123217; PubMed=1991729;
RA   Parales J. Jr., Greenberg E.P.;
RT   "N-terminal amino acid sequences and amino acid compositions of the
RT   Spirochaeta aurantia flagellar filament polypeptides.";
RL   J. Bacteriol. 173:1357-1359(1991).
CC   -!- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC   -!- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF
CC       REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS ONE OR ALL
CC       FIVE ANTIGENICALLY RELATED POLYPEPTIDES.
CC   -!- SUBCELLULAR LOCATION: Periplasmic flagellum.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M24450; AAA[illegible]27.1; -.
DR   PIR; A32814; A32814.
KW   Flagella; Periplasmic; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    337       FLAGELLAR FILAMENT OUTER LAYER PROTEIN.
SQ   SEQUENCE   337 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             [illegible];  Score [illegible];  DB [illegible];  Length 337;
  Best Local Similarity   30.4%;  Pred. No. 1.5e-22;
  Matches  [illegible];  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy          1 M[illegible]NYVD[illegible] 60
Db          1 MEE[illegible]NIGIHAPIIIE 47

Qy         61 [illegible] 120
Db         48 [illegible]AAPYGR 106

Qy        121 [illegible]AM[illegible] 167
Db        107 [illegible] 166

Qy        168 [illegible] 227
Db        167 [illegible] 226

Qy        228 [illegible] 287
Db        227 [illegible] 287

Qy        288 [illegible]KAESEEK 306
Db        288 [illegible]VLRSLEEKK 314


RESULT 3
FLAA_TREPA
ID   FLAA_TREPA     STANDARD;      PRT;   350 AA.
AC   P18193;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Flagellar filament outer layer protein precursor (Sheath protein).
GN   FLAA OR TP0249.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90[illegible]; PubMed=2194955;
RA   Isaacs R.D., Radolf J.D.;
RT   "Expression in Escherichia coli of the 37-kilodalton endoflagellar
RT   sheath protein of Treponema pallidum by use of the polymerase chain
RT   reaction and a T7 expression system.";
RL   Infect. Immun. 58:2025-2034(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
```

[illegible]

```
ID   FLAI_TREHY     STANDARD;      PRT; [illegible] AA.
AC   P32[illegible]; P80[illegible];
DT   [illegible] (Rel. [illegible], Created)
DT   [illegible] (Rel. [illegible], Last sequence update)
DT   [illegible] 2001 (Rel. 4[illegible], Last annotation update)
DE   Flagellar filament [illegible] layer protein [illegible] ([illegible] kDa [illegible]
DE   protein).
GN   FLAA1 OR FLAA.
OS   Treponema hyodysenteriae (Serpulina hyodysenteriae).
OC   Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
OX   NCBI_TaxID=[illegible];
RN   [1]
RP   SEQUENCE FROM N.A. AND SEQUENCE [illegible].
RC   STRAIN [illegible];
RX   MEDLINE [illegible]; PubMed [illegible];
RA   Koopman M.B.H., [illegible] Leeuw [illegible], van der [illegible] B.A.M., Kusters [illegible];
RT   "Cloning and DNA sequence analysis of a Serpulina (Treponema)
RT   hyodysenteriae gene encoding a periplasmic flagellar sheath
RT   protein.";
RL   Infect. Immun. [illegible] (19[illegible]).
RN   [2]
RP   SEQUENCE OF [illegible].
RC   STRAIN [illegible];
RX   MEDLINE [illegible]; PubMed [illegible];
RA   Koopman M.B.H., Baats E., van Vorstenbosch [illegible], [illegible] A.B.V.,
RA   van der Zeijst B.A.M., Kusters J.G.;
RT   "The periplasmic flagella of Serpulina (Treponema) hyodysenteriae are
RT   composed of two sheath proteins and three core proteins.";
RL   J. Gen. Microbiol. 138:26[illegible] (1992).
CC   -!- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC   -!- SUBUNIT: THE FLAGELLUM [illegible] AN OUTER LAYER COMPOSED OF TWO
CC       SHEATH PROTEINS, FLAA1 (44 KDA) AND FLAA2 ([illegible] KDA) AROUND A CORE
CC       THAT CONTAINS THREE PROTEINS FLAB1 ([illegible] KDA), FLAB2 ([illegible] KDA), AND
CC       FLAB3 ([illegible] KDA).
CC   -!- SUBCELLULAR LOCATION: Periplasmic flagellum.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X6[illegible]; CAA44[illegible].1; -.
DR   EMBL; A24[illegible]; CAA[illegible].1; -.
DR   PIR; A4[illegible]24; A4[illegible].
KW   Flagellar; Signal; Periplasmic.
FT   SIGNAL        1     [illegible]
FT   CHAIN         [illegible]   [illegible]      FLAGELLAR FILAMENT OUTER LAYER PROTEIN
FT                                  FLAA1.
FT   CONFLICT      [illegible]   [illegible]      [illegible] (IN REF. [illegible]).
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches [illegible];  Conservative [illegible];  Mismatches [illegible];  Indels [illegible];  Gaps [illegible];

Qy   [illegible]
Db     4 LEVVLISIFIAASAY[illegible]NS[illegible] [illegible]
Qy    [illegible] IVKEIDMVV[illegible]
Db    [illegible]
Qy    [illegible]
Db    [illegible] AAWNSYAIVK[illegible]
Qy   187 EDMNR[illegible]
Db    [illegible]
```

```
Qy     247 SKSHSSFV[illegible]VYE[illegible]K[illegible] 2[illegible]

Db     223 Y[illegible]K      [illegible] 2[illegible]

Qy     3[illegible]7 RVLKLREKISIAE 3[illegible]9

Db     277   ARIREQAELRD 287


RESULT 5
YCXD_BACSU
ID   YCXD_BACSU     STANDARD;      PRT;   444 AA.
AC   [illegible]8792;
DT   01 FEB 1995 (Rel. 31, Created)
DT   01 FEB 1995 (Rel. 31, Last sequence update)
DT   15-JUN 2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ycxD ([illegible]).
GN   YCXD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID 1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN [illegible];
RX   MEDLINE 93360813; PubMed [illegible].
RA   Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA   Venema G., van Sinderen D.;
RT   "Sequence and analysis of the genetic locus responsible for surfactin
RT   synthesis in Bacillus subtilis.";
RL   Mol. Microbiol. 8:821 831(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN 168.
RX   MEDLINE 97124189; PubMed 8969[illegible]2;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees 36 degrees region of the Bacillus subtilis chromosome:
RT   determination of the sequence of a 146 kb segment and identification
RT   of 113 genes.";
RL   Microbiology [illegible](1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN 168;
RX   MEDLINE [illegible];
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   [illegible]
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   [illegible]
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., [illegible]
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249 256(1997).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (POTENTIAL).
CC   -!- SIMILARITY: IN THE C TERMINAL, TO AMINOTRANSFERASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://[illegible]
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X70356; CAA49821.1; -.
DR   EMBL; D50453; BAA0899[illegible].1; -.
DR   EMBL; Z99105; CAB12[illegible].1; -.
DR   EMBL; Z99106; CAB12164.1; -.
DR   SubtiList; BG1017[illegible]; ycxD.
DR   InterPro; IPR000524; HTH_GntR.
DR   Pfam; PF00392; gntR; 1.
DR   PRINTS; PR00035; HTHGNTR.
DR   SMART; SM00345; HTH_GNTR; 1.
KW   Hypothetical protein; Transferase; Aminotransferase;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     28[illegible]    28[illegible]       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   444 AA;  50755 MW;  [illegible] CRC64;

  Query Match             6.3%;  Score 109;  DB 1;  Length 444;
  Best Local Similarity   21.6%;  Pred. No. [illegible];
  Matches   75;  Conservative   42;  Mismatches  126;  Indels  104;  Gaps   1[illegible];

Qy      2[illegible] SKRAEPGELVLDFAELARDPSSIRLDLINYVD          YVYS[illegible]AS[illegible]  [illegible]

Db      7[illegible] [illegible]KSGQIGI  IDFAISAFD  [illegible]NKA[illegible]YKN[illegible]

Qy     [illegible]   VKPLDMVV[illegible]NNWSVL[illegible]   [illegible]AS[illegible]AYVEN [illegible]

Db     12[illegible] [illegible]MVF[illegible] 1[illegible]

Qy     1[illegible] [illegible]VVAPAVVE[illegible]KFYA[illegible]NQFLSK[illegible]

Db     1[illegible] [illegible]AKVER[illegible]MPRFHNP  [illegible]

Qy     1[illegible] [illegible]MN[illegible]

Db     2[illegible] [illegible]

Qy     2[illegible] [illegible]VENF[illegible]VSI[illegible]

Db     2[illegible] [illegible]EMME[illegible]AAV[illegible]

Qy     281 [illegible]SVEKVY[illegible]SIAPDSF 322

Db     3[illegible] [illegible]AALE[illegible]YLK[illegible] 3[illegible]8


RESULT 6
Y686_METJA
ID   Y686_METJA     STANDARD;      PRT;   [illegible] AA.
AC   Q58099;
DT   01 NOV 1997 (Rel. 35, Created)
DT   01 NOV 1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0686.
GN   MJ0686.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID 2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN [illegible] 10067;
RX   MEDLINE 96337999; PubMed 8688087;
```

```
[illegible] Bult C.J. [illegible]
[illegible] Weinstock K.G. [illegible] Glodek A.,
[illegible] Nguyen D.,
[illegible] Hanna M.C.,
[illegible] Borodovsky M.,
[illegible] Methanococcus

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

Qy [illegible] AMIMGI[illegible] IINKIMKEIKV [illegible]
Db [illegible] KMMKIEKIV [illegible] VVENIKI [illegible]
Qy [illegible] YIENISSKIIKI [illegible]
Db [illegible] KAHIIEKEVLKI [illegible]
```

```
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., [illegible] Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., [illegible] Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:[illegible](1996).
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein [illegible].
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U67[illegible]; AAB9[illegible].
DR   TIGR; MJ12[illegible]; -.
DR   InterPro; IPR0[illegible]; V_ATPase_sub[illegible].
DR   Pfam; PF014[illegible]; V_ATPase_sub_a; 1.
KW   Hydrolase; Hydrogen ion transport; Transmembrane; [illegible]
FT   TRANSMEM    [illegible]    [illegible]       POTENTIAL.
FT   TRANSMEM    [illegible]    [illegible]       POTENTIAL.
FT   TRANSMEM    [illegible]    [illegible]       POTENTIAL.
FT   TRANSMEM    [illegible]    [illegible]       POTENTIAL.
FT   TRANSMEM    [illegible]    [illegible]       POTENTIAL.
FT   TRANSMEM    [illegible]    [illegible]       POTENTIAL.
FT   TRANSMEM    [illegible]    [illegible]       POTENTIAL.
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match              [illegible]%;  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity    [illegible]%;  Pred. No. [illegible];
  Matches   [illegible];  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy      [illegible] IIVKIEIMVV [illegible] NNWSV[illegible] SARI[illegible] AYVEN[illegible] VVAI[illegible] AVVE[illegible] YA [illegible] VEVI [illegible]

Db      [illegible] IIVEI [illegible] WKI[illegible] SSSAI YVEN [illegible] MIE [illegible]

Qy      127 [illegible] SYS[illegible] SSAMI [illegible] MEIEKI[illegible] [illegible]

Db      [illegible] MESSVQ[illegible] KII[illegible] KKKV[illegible] VII[illegible] AIEVIN[illegible]

Qy      162 NIKI-MKPIKVEV [illegible] IIPVIPII MN[illegible] MIYAY [illegible] WAI [illegible]

Db      137 NKKSKILQIK[illegible] IYIK[illegible] IEFIIKYI [illegible] IYVE[illegible] VIKIK [illegible] AIII [illegible]

Qy      211 WQNENYIEN[illegible] VENY[illegible] KMEIKAIEV[illegible] KVEN [illegible] IYVKI[illegible] EVIYI [illegible]

Db      189 KVAI[illegible] KKIEVI IVEVI[illegible] IKIINV [illegible] EYII [illegible]

Qy      2[illegible] KISVSI[illegible] ISVEKVYI[illegible] AEIIEKVIKI[illegible]

Db      24[illegible] [illegible] SIV [illegible] SEAI[illegible] SEIKAIESEN[illegible] KAIA[illegible] EWIKE[illegible] AVVE [illegible]

Qy      [illegible] AI[illegible] N [illegible] EVIKIIV[illegible]

Db      29[illegible] KAEQAYS[illegible] IKIIEIYYIEAWVIAEIAEKAK[illegible] IENIAI[illegible] AEVI[illegible] IEK[illegible]


RESULT 8
VWBA_BACSU
ID   VWBA_BACSU     STANDARD;      PRT;   [illegible] AA.
AC   [illegible]2113[illegible];
DT   01-JUL-19[illegible] (Rel. [illegible], Created)
DT   01-JUL-19[illegible] (Rel. 2[illegible], Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Amidase enhancer precursor (M[illegible]).
GN   LYTB [illegible] VWBA.
```

```
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID 1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN 168;
RX   MEDLINE 93018998; PubMed 1357079;
RA   Lazarevic V., Margot P., Soldo B., Karamata D.;
RT   "Sequencing and analysis of the Bacillus subtilis lytRABC divergon: a
RT   regulatory unit encompassing the structural genes of the N
RT   acetylmuramoyl-L alanine amidase and its modifier.";
RL   J. Gen. Microbiol. 138:1949-1961(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE [illegible]; PubMed [illegible];
RA   Kuroda A., Rashid H.M., Sekiguchi J.;
RT   "Molecular cloning and sequencing of the upstream region of the major
RT   Bacillus subtilis autolysin gene: a modifier protein exhibiting
RT   sequence homology to the major autolysin and the spoIID product.",
RL   J. Gen. Microbiol. 138:1067-1076(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN 168;
RX   MEDLINE 98044033; PubMed 9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: POSSIBLY INVOLVED IN CELL WALL METABOLISM DURING SPORE
CC       FORMATION. ENHANCES THE AMIDASE ACTIVITY APPROXIMATELY THREEFOLD.
CC   -!- SUBCELLULAR LOCATION: Membrane bound.
CC   -!- SIMILARITY: WITH THE N TERMINAL OF LYTC AT ITS N TERMINAL AND THE
CC       B.SUBTILIS SPOIID GENE AT ITS C TERMINAL.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M87645; AAA22579.1; -.
DR   EMBL; D10388; BAA01224.1; -.
DR   EMBL; Z99122; CAB15588.1; -.
DR   PIR; C47679; C47679.
DR   SubtiList; BG10406; lytB.
KW   Membrane; Signal; Cell wall; Repeat; Complete proteome.
FT   SIGNAL        1     25
FT   CHAIN        26    705       AMIDASE ENHANCER.
FT   DOMAIN       64    349       3 X TANDEM REPEATS
FT   REPEAT       64    161       1.
FT   REPEAT      162    251       2.
FT   REPEAT      252    349       3.
SQ   SEQUENCE   705 AA;  76726 MW;  ATA2448[illegible]8015[illegible]8 CRC64;

  Query Match             6.1%;  Score 105.5;  DB 1;  Length 705;
  Best Local Similarity   25.1%;  Pred. No. 2.9;
  Matches   76;  Conservative   32;  Mismatches  125;  Indels   70;  Gaps   14;

Qy         45 [illegible] 103
Db        252 RIPGSTRYETANIIEQLK       LEADKVVMINDIFYADVI TAP    AYKKNSQ[illegible] 302

Qy        104          APAVVKSHSK    RYA[illegible]VEVL[illegible]AMIMP[illegible]NQF 154
Db        303 FVEQ[illegible]VPAAAK[illegible]KAIYAY[illegible]       [illegible]AEVEK[illegible]N 355

Qy        155 I       [illegible]NIKI    MK[illegible]FV[illegible]VY[illegible] 207
Db        356 [illegible]NLENIKTY[illegible]      [illegible]KVNES      402

Qy        208 [illegible]WSNPNYIPNIGSR:       [illegible]EVEN[illegible] 261
Db        403       [illegible]KV[illegible]VNEN[illegible]NEM[illegible]KA[illegible]           454

Qy        262 [illegible] 321
Db        455                     [illegible]TIVEI[illegible]NKAVE[illegible]HAAYS 500

Qy        322 FQN 324
                |
Db        501 SQN 503


RESULT 9
ITH1_PIG
ID   ITH1_PIG       STANDARD;      PRT;   902 AA.
AC   Q29052; Q02669;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Inter-alpha-trypsin inhibitor heavy chain H1 precursor (ITI heavy
DE   chain H1).
GN   ITIH1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID 9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Suzuki H., Hamasima N., Fimori M., [illegible] A., Yasue H.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 124-902 FROM N.A.
RC   TISSUE=Liver;
RA   Gebhard W.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A
CC       BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,
CC       INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE
CC       LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE
CC       ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: I ALPHA I PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM
CC       ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN,
CC       BIKUNIN. INTER ALPHA INHIBITOR (I ALPHA I) IS COMPOSED OF H1, H2
CC       AND BIKUNIN. INTER ALPHA LIKE INHIBITOR (I ALPHA LI) OF H2 AND
```

AND BIKUNIN (BY

HEAVY CHAIN

(BY SIMILARITY).

KAPVQA 521

KARVQA 521

DB [illegible] KIFIVM [illegible] MKKL [illegible] FWAE [illegible]

RESULT 1
XYLB_BACSU
ID XYLB_BACSU STANDARD; PRT; [illegible] AA.
AC P39211;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Xylulose kinase (EC 2.7.1.17) (Xylulokinase).
GN XYLB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Borchert S., Klein C., Piksa B., Hammelmann M., Entian K.-D.;
RT "Sequencing of a [illegible] kb region of the Bacillus subtilis genome downstream of spoVI.";
RL Submitted (FEB-199[illegible]) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044[illegible]; PubMed=[illegible];
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., [illegible]
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., [illegible]
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., [illegible]
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., [illegible]
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., [illegible]
RA Denizot F., Devine K.M., Duesterhoeft A., Ehrlich S.D., Emmerson P.T., [illegible]
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., [illegible]
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., [illegible]
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., [illegible]
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., [illegible]
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., [illegible]
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., [illegible]
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M., [illegible]
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., [illegible]
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., [illegible]
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., [illegible]
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., [illegible]
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., [illegible]
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., [illegible]
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., [illegible]
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., [illegible]
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., [illegible]
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., [illegible]
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., [illegible]
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., [illegible]
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T., [illegible]
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., [illegible]
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-[illegible] FROM N.A.
RX MEDLINE=85[illegible]; PubMed=[illegible];
RA Wilhelm M., Hollenberg C.P.;
RT "Nucleotide sequence of the Bacillus subtilis xylose isomerase gene: extensive homology between the Bacillus and Escherichia coli enzyme.";
RL Nucleic Acids Res. 13:5717-5722(1985).
CC -!- CATALYTIC ACTIVITY: ATP + D-xylulose = ADP + D-xylulose 5-phosphate.
CC -!- SIMILARITY: BELONGS TO THE FGGY KINASE / [illegible]N KINASE / XYLULOKINASE / XYLB KINASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U66480; AAB41094.1; -.
DR   EMBL; Z99113; CAB13645.1; -.
DR   EMBL; X52795; CAA26563.1; -.
DR   SubtiList; BG10807; xylB.
DR   InterPro; IPR000577; FGGY kin.
DR   Pfam; PF00370; FGGY; 1.
DR   Pfam; PF02782; FGGY_C; 1.
DR   TIGRFAMs; TIGR01312; XylB; 1.
DR   PROSITE; PS00445; FGGY_KINASES_2; 1.
DR   PROSITE; PS00933; FGGY_KINASES_1; 1.
KW   Transferase; Kinase; Xylose metabolism; Complete proteome.
SQ   SEQUENCE   499 AA;  55417 MW;  [illegible] CRC64;

  Query Match             6.1%;  Score 104;  DB 1;  Length 499;
  Best Local Similarity   20.9%;  Pred. No. 2.3;
  Matches   68;  Conservative   57;  Mismatches  135;  Indels   66;  Gaps   14;

Qy         30 [illegible] 81
Db        141 [illegible] 199
Qy         82 [illegible] 141
Db        200 [illegible] 229
Qy        142 [illegible]
Db        230 [illegible]
Qy        194 [illegible] 2??
Db        295 [illegible]
Qy        2?? [illegible]
Db        34? [illegible]
Qy        2?? [illegible]
Db        4?? [illegible]

RESULT 11
HS71_LYCES
ID   HS71_LYCES     STANDARD;      PRT;   650 AA.
AC   P24629;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Heat shock cognate 70 kDa protein 1.
GN   HSC-1.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. VF36; TISSUE=Pistil;
RX   MEDLINE=91370894; PubMed=1893113;
RA   Lin T.Y., Duck N.B., Winter J., Folk W.R.;
RT   "Sequences of two hsc 70 cDNAs from Lycopersicon esculentum.";
RL   Plant Mol. Biol. 16:475-478(1991).
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X54029; CAA37975.1; -.
DR   PIR; S14949; S14949.
DR   HSSP; P19120; 1HPM.
DR   InterPro; IPR001023; Hsp70.
DR   Pfam; PF00012; HSP70; 1.
DR   PRINTS; PR00301; HEATSHOCK70.
DR   ProDom; PD000089; Hsp70; 1.
DR   PROSITE; PS00297; HSP70_1; 1.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   ATP-binding; Heat shock; Multigene family.
SQ   SEQUENCE   650 AA;  71287 MW;  [illegible] CRC64;

  Query Match             [illegible];  Score [illegible];  DB 1;  Length [illegible];
  Best Local Similarity   18.8%;  Pred. No. [illegible];
  Matches   68;  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible];

Qy         [illegible]
Db         6? [illegible] 1??
Qy         1?? [illegible] 142
Db         10? [illegible]
Qy         1?? [illegible]
Db         1?? [illegible]
Qy         1?? [illegible]
Db         2?? [illegible]
Qy         2?? [illegible]
Db         27? [illegible]
Qy         2?? [illegible]
Db         ??? [illegible]
Qy        337 S 337
Db        375 A 375

RESULT 12
ITIH_MESAU
ID   ITIH_MESAU     STANDARD;      PRT;   [illegible] AA.
AC   P97278;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Inter-alpha-trypsin inhibitor heavy chain H1 precursor (ITI heavy
DE   chain H1) (HC1).
GN   ITIH1.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97420688; PubMed=9276673;
RA   Nakatani T., Suzuki Y., Yamamoto T., Sinohara H.;
RT   "Molecular cloning and sequencing of cDNAs encoding three heavy-chain
RT   precursors of the inter-alpha-trypsin inhibitor in Syrian hamster:
RT   implications for the evolution of the inter-alpha-trypsin inhibitor
```

[illegible]

Qy 274 SIDSDIDSESVEKVYEISG IESLEKIKAHIIEKEVIKIEIKI [illegible]

Db [illegible] KAICIVEFEMKKIEEKGHMIFNHVEIIWAY [illegible] AFIMKM [illegible]

RESULT 13
CGAA_CLOBI
ID CGAA_CLOBI STANDARD; PRT; [illegible] AA.
AC Q45882;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidial crystal like protein cry16Aa (Insecticidal toxin
DE CryXVIA(a)) (Cbm71 mosquitocidal toxin).
GN CRY16AA OR CRYXVIA(A) OR CBM71.
OS Clostridium bifermentans.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; [illegible]
OC Clostridium.
OX NCBI_TaxID [illegible];
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE [illegible] AND [illegible]
RC STRAIN [illegible] / subsp. Malaysia;
RX MEDLINE [illegible]; PubMed [illegible];
RA Barloy F., [illegible] A., [illegible] M.
RT "Cloning and expression of the first [illegible] toxin gene from
RT Clostridium bifermentans subsp. malaysia, encoding a new [illegible]
RT protein with [illegible] Bacillus thuringiensis [illegible] toxins."
RL J. Bacteriol. [illegible]
CC -!- FUNCTION: TOXIN ACTIVE ON MOSQUITO LARVAE [illegible]
CC AEGYPTI, CULEX PIPIENS, AND ANOPHELES STEPHENSI.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DEVELOPMENTAL STAGE: PRODUCED DURING SPORULATION. THE EXPRESSION
CC DECREASES CONSIDERABLY WITH CELL LYSIS.
CC -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X94146; CAA63860.1; -.
DR InterPro; IPR[illegible]; Endotoxin.
DR Pfam; PF[illegible]; [illegible]
KW Toxin; Sporulation.
FT DOMAIN [illegible]
FT CONFLICT [illegible] (IN REF. [illegible] AA SEQUENCE).
SQ SEQUENCE [illegible] AA; [illegible] MW; [illegible] CRC64;

Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Qy 42 AELARIESGI [illegible] NYVIYVY [illegible] VKIL MVVI [illegible] NNNAV [illegible] AE [illegible] AYVKN [illegible]

Db 209 CNIKLIINNSNI [illegible] Y [illegible] M [illegible] M [illegible] KVY [illegible] K [illegible]

Qy [illegible] VVAFAVVKIEKIYA [illegible] VEVIII [illegible]

Db 250 [illegible] FINMQII SKVYS [illegible] VYEKMEN [illegible] VINK [illegible] NKE [illegible]

Qy [illegible] K [illegible] NIKIMKIIKVAVY [illegible] MN [illegible] NEVAY [illegible]

Db [illegible] NGHVNKGV [illegible] NINW [illegible] NY [illegible] NNNI [illegible] NNNE [illegible] VK [illegible]

```
Qy     2?? ILKF    KGWA[illegible]IW[illegible]N[illegible]N          YI[illegible]N[illegible]KDDV[illegible]NY[illegible]LAS[illegible] KMK    240
Db     366 SIF[illegible]K[illegible]EWI[illegible]N[illegible]YL[illegible]N[illegible]YIHS[illegible]MILANN[illegible]INI 425

Qy     241         FEA[illegible]VK[illegible]KVKNE[illegible]YVK[illegible]    YDK[illegible]V 27?
Db     426 DIFHSY[illegible]WI[illegible]NY[illegible]K[illegible]N[illegible]LVK[illegible]K[illegible]HYSEISV 472


RESULT 14
YF?4 METJA
ID   YF?4 METJA      STANDARD;      PRT;   622 AA.
AC   Q58949;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ1554 precursor.
GN   MJ1554.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: TO A.FULGIDUS AF0817.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U67596; AAB99574.1; -.
DR   TIGR; MJ1554; -.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    622       HYPOTHETICAL PROTEIN MJ1554.
SQ   SEQUENCE   622 AA;  72282 MW;  [illegible]AA[illegible] CRC64;

  Query Match             5.9%;  Score 101.5;  DB 1;  Length 622;
  Best Local Similarity   2?.?%;  Pred. No. 4.?;
  Matches   ??;  Conservative   74;  Mismatches  149;  Indels   ??;  Gaps   ??;

Qy       1 MEEKAK[illegible]A[illegible]FA[illegible]NYV 59
Db       1 MKIKAVAI[illegible]VENEKPIK[illegible]N[illegible]FKL[illegible]VN[illegible]NFEEEKNTV 54

Qy      6? [illegible]YVY[illegible]NNW[illegible]AYVEN[illegible]VVAPAVV[illegible]KR 114
Db      55 ENS[illegible]NY[illegible]VKS[illegible]N    VEIS[illegible]ERFSKINV[illegible]VK[illegible]V 110

Qy     115 YAG[illegible]LGVRVLFPSYSQS?AMIMPPPKIPPYS       CFSCNQP[illegible]KCLIDNIKTMKE 168
Db     111 INGN[illegible]          A[illegible]NEIYL[illegible]EYAK[illegible]GYEYLIN [illegible]NILIVIEW 153

Qy     1?? IEVAVY[illegible]LEVLFL[illegible]YAYSM    [illegible]KF  KGWAL  LIWGN        21?
Db     1?? NKIISYNV[illegible]IWQM[illegible]VVRKN[illegible]IVWNNYK[illegible]YD 217
```

```
Qy     214 ENYIENI[illegible]VENYILA[illegible]EKAFEV[illegible]       K[illegible]FVKNE[illegible]YVKIE 264
Db     218 KYYIFELI[illegible]YGMDFD[illegible]     [illegible]NIKGQ[illegible]VEN[illegible]IAIV[illegible]NTK[illegible]MGKNNI[illegible]AYNI 2??

Qy     265 KVLYDKLSVS[illegible]DSDIDSESVEKVYEIS[illegible]ESLKKLKAHF[illegible]KEVI[illegible]EKISIA[illegible]N 3?4
Db     2?? [illegible]NEEFLM[illegible]    NE[illegible]A[illegible]VAIE[illegible]NQ[illegible]    VEIIE [illegible]EE 32?

Qy     32? IVEKIESEK 333
Db     328 YLSS[illegible]PSEK 33?


RESULT 15
HS7?_TRYBB
ID   HS7C_TRYBB     STANDARD;      PRT;   676 AA.
AC   P20030;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-MAY-1992 (Rel. 22, Last annotation update)
DE   Heat shock cognate HSP70 protein.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90377290; PubMed=2398917;
RA   Lee M.G.-S., Polvere R.I., van der Ploeg L.H.T.;
RT   "Evidence for segmental gene conversion between a cognate hsp70 gene
RT   and the temperature-sensitively transcribed hsp70 genes of
RT   Trypanosoma brucei.";
RL   Mol. Biochem. Parasitol. 41:213-220(1990).
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M32139; AAA30208.1; -.
DR   PIR; [illegible]; A45515.
DR   HSSP; P19120; 3HSC.
DR   InterPro; IPR001023; Hsp70.
DR   Pfam; PF00012; HSP70; 1.
DR   ProDom; PD000089; Hsp70; 1.
DR   PROSITE; PS00297; HSP70_1; 1.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   ATP-binding; Heat shock; Multigene family.
SQ   SEQUENCE   676 AA;  7?094 MW;  AB67B7831246785E CRC64;

  Query Match             5.9%;  Score 101.?;  DB 1;  Length 676;
  Best Local Similarity   2?.?%;  Pred. No. ?.?;
  Matches   73;  Conservative   44;  Mismatches  11?;  Indels   ??;  Gaps   ??;

Qy      63 YSGASGIVKPEDMVVDLG              INNWSVLLIPSARLQAYVKNSVVAP 105
Db       3 [illegible]EGAIGI      [illegible]VGVWQNERVEIIAN[illegible]NRT[illegible]  YVAFVNNEVIV[illegible] ?.

Qy     10? AVVKSESKRYA[illegible]V RVLFPSYSQS    SAMIMPPPKIPPYSQES?N     [illegible] 15?
Db      5? DAAKSHAAF[illegible]NVII[illegible]AKEL[illegible]VV[illegible]MPEWEFKV    [illegible]AVMEVEHRG 110

Qy     157 FGLIDNIKTMKEIEVAVY[illegible]LEVLFL[illegible]YAY[illegible]KF[illegible]WAL[illegible]NY 2??
Db     111 EGML[illegible]  LQPEQISARVLAY   LKS?AESYL[illegible]VAKAVVIV           FAY 15?

Qy     217 IPNI[illegible]         IVENYILA[illegible]EKAFEV[illegible]EVKNE[illegible]YVKIELEV 26?
```

[illegible] Searchads
[illegible] cell updates/sec

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	[illegible]	4.7	[illegible]	24	ABI[illegible]	Human [illegible]
13	76.8	4.6	[illegible]	24	ABI[illegible]	Human immune [illegible]
14	76.8	4.6	[illegible]	24	AAS[illegible]	Human [illegible]
[illegible]	[illegible]	4.[illegible]	[illegible]	24	AB[illegible]	Human immune [illegible]
[illegible]	76.4	4.[illegible]	[illegible]	[illegible]	AAS[illegible]	[illegible]
17	[illegible]	4.[illegible]	[illegible]	24	ABK[illegible]	[illegible]
18	75.4	4.[illegible]	[illegible]	24	ABI[illegible]	Human immune [illegible]
19	74.8	4.5	[illegible]	24	ABI[illegible]	Chemically treated
20	74.8	4.5	[illegible]	24	ABI[illegible]	Human immune [illegible]
21	74.8	4.5	[illegible]	24	ABI[illegible]	Human [illegible]
22	74.8	4.5	[illegible]	24	ABK[illegible]	[illegible]
23	74.[illegible]	4.5	[illegible]	22	AAS[illegible]	[illegible]
24	73.8	4.4	[illegible]	24	ABI[illegible]	Human immune [illegible]
25	73.6	4.4	[illegible]	22	AAI[illegible]	Human [illegible]
26	73.4	4.4	[illegible]	24	ABI[illegible]	Human immune [illegible]
27	73.4	4.4	[illegible]	24	ABI[illegible]	Human immune [illegible]
28	73.4	4.4	[illegible]	24	AB[illegible]	Human [illegible]
29	72.6	4.4	[illegible]	24	ABI[illegible]	Human immune [illegible]
30	72.6	4.4	[illegible]	24	ABK[illegible]	DNA transcription [illegible]
31	72.4	4.4	[illegible]	24	ABI[illegible]	Human immune [illegible]
32	71.8	4.3	[illegible]	22	AAS[illegible]	[illegible]
33	71.8	4.3	[illegible]	24	ABI[illegible]	Human immune [illegible]
34	71.[illegible]	4.3	[illegible]	24	ABI[illegible]	Human immune [illegible]
35	71.4	4.3	[illegible]	22	AAB[illegible]	Human [illegible]
36	71	4.3	[illegible]	24	ABI[illegible]	Chemically treated
37	71	4.3	[illegible]	24	ABI[illegible]	Human [illegible]
38	[illegible]	4.3	[illegible]	22	AAS[illegible]	[illegible]
39	[illegible]	4.3	[illegible]	24	ABI[illegible]	Human immune [illegible]
40	[illegible]	4.2	[illegible]	[illegible]	AAS[illegible]	[illegible]
41	[illegible]	4.2	[illegible]	24	ABI[illegible]	Human immune [illegible]
42	70.2	4.2	14319	22	AAS[illegible]	Tumour suppressor
43	70	4.2	[illegible]	24	ABI[illegible]	Human [illegible]
44	70	4.2	[illegible]	24	ABI[illegible]	Human immune [illegible]
45	[illegible]	4.2	[illegible]	24	ABI[illegible]	Human immune [illegible]

ALIGNMENTS

RESULT 1
AAX8[illegible]
ID AAX8[illegible] standard; DNA; [illegible] BP.
XX
AC AAX8[illegible];
XX
DT [illegible] (first entry)
XX
DE Nucleic acid sequence encoding the [illegible]A gene product.
XX
KW [illegible]; [illegible]A; Lyme disease; [illegible]
XX
OS Synthetic.
OS Borrelia burgdorferi.
XX
FH Key Location/Qualifiers
FT CDS [illegible]
FT /*tag= [illegible]
FT /product= "[illegible]/[illegible]A [illegible]"
FT sig_peptide [illegible]
FT /*tag= b
FT mat_peptide [illegible]..14[illegible]
FT /*tag= [illegible]
XX
PN W[illegible] A1.
XX
PD 15-JUL-[illegible].
XX
PF [illegible]-JAN-[illegible]; [illegible]W[illegible].
XX
PR [illegible]-JAN-[illegible]; [illegible].
XX

```
PA   (INNE-) BIOMERIEUX INC.
XX
PI   Gilmore RD,  Johnson BJB;
XX
DR   WPI; 1999-430469/36.
DR   P-PSDB; AAY28460.
XX
PT   Recombinant FlaA/P37 protein antigen, useful in improved detection
PT   of Borrelia burgdorferi
XX
PS   Disclosure; Page 28; 43pp; English.
XX
CC   The sequence encodes the recombinant protein antigen P37/FlaA. The
CC   P37 protein product can be used in an assay for the detection of Borrelia
CC   burgdorferi infection (the causative agent of Lyme disease). FlaA is
CC   an outer sheath protein of the periplasmic flagella of B. burgdorferi.
CC   FlaA is a prominent antigen in the early humoral immune response to
CC   Borrelia burgdorferi infection, and significantly suitable for use in
CC   improved serologic tests for exposure to Lyme disease spirochetes. The
CC   detection of FlaA can augment the set of recombinant molecules that are
CC   recognized early in the course of disease and contribute to improved
CC   sensitivity of early testing for Lyme disease.
XX
SQ   Sequence 1655 BP; 560 A; 182 C; 308 G; 605 T; 0 other;

  Query Match             98.4%;  Score 1637;  DB 20;  Length 1655;
  Best Local Similarity   99.5%;  Pred. No. 7.4e-262;
  Matches 1655;  Conservative    0;  Mismatches    0;  Indels    8;  Gaps    1;

Qy          1 ATGATAATCTTTTTCAAAAAAGGTTTTTATTTTCATTCTAGCAAGGATTTGTTGCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATAATCTTTTTCAAAAAAGGTTTTTATTTTCATTCTAGCAAGGATTTGTTGCTA 60

Qy         61 ATTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAG 120

Qy        121 ATTTTCTTTTAGGAAGCAATGAGGGTTTTCTTTTGGTTTTTATTAAGTGATTCAAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTTCTTTTAGGAAGCAATGAGGGTTTTCTTTTGGTTTTTATTAAGTGATTCAAGAT 180

Qy        181 TTTTATATTCTTTTTAAAGAATGGAGTTTATTATGTAAATCTTTCAAGAGAATTTTATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTATATTCTTTTTAAAGAATGGAGTTTATTATGTAAATCTTTCAAGAGAATTTTATG 240

Qy        241 ATTCTTTTAATAATGGTGATTATAATGAATCTTTTGATGTTAAGGTCAATCTTTTTGCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCTTTTAATAATGGTGATTATAATGAATCTTTTGATGTTAAGGTCAATCTTTTTGCTA 300

Qy        301 TTTCTTTAATAAAAACAATGTGCTTTAACTATCTTGTTAAGATAAAAAAGATTATTATTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCTTTAATAAAAACAATGTGCTTTAACTATCTTGTTAAGATAAAAAAGATTATTATTC 360

Qy        361 TTGTTGAAGGGTGTATCTTAAAGGAGCAAAGTTGATAAATTAAATTTTACTAATAAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGTTGAAGGGTGTATCTTAAAGGAGCAAAGTTGATAAATTAAATTTTACTAATAAAAAT 420

Qy        421 AATTAAAAAACGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTACATGAAAAG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AATTAAAAAACGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTACATGAAAAG 480

Qy        481 GAAAGCTAAAAGTATTTTATTTTTTTTATTATCCACTGTTCTTTTTGCTCAAGAGACTGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAAGCTAAAAGTATTTTATTTTTTTTATTATCCACTGTTCTTTTTGCTCAAGAGACTGA 540

Qy        541 TGGATTAGCAGAGGGTTCTAAAAGGGCAGAGCTTGGAGAATTAGTCTTAGATTTTGCCGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGATTAGCAGAGGGTTCTAAAAGGGCAGAGCTTGGAGAATTAGTCTTAGATTTTGCCGA 600

Qy        601 GCTTGCAAGAGATCCAAGTTCAACTAGACTTGATCTTACAAATTATGTTGATTATGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCTTGCAAGAGATCCAAGTTCAACTAGACTTGATCTTACAAATTATGTTGATTATGTATA 660

Qy        661 TCGGGGGCTTCTGGTATTGTTAAGCCGGAAGATATGGTTGGAGATTTGGGATAAATAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCGGGGGCTTCTGGTATTGTTAAGCCGGAAGATATGGTTGGAGATTTGGGATAAATAA 720

Qy        721 TGCAGCGTTTTACTTACTCCTTCTGCAAGGTTGCAGGTTACGTTAAAAATTCAGTTGT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCAGCGTTTTACTTACTCCTTCTGCAAGGTTGCAGGTTACGTTAAAAATTCAGTTGT 780

Qy        781 GCCCCGCTGTTGTTAAGAGTGAGTCAAAAAGGTACGCAGGTGATATCATTTTGGGTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCCCGCTGTTGTTAAGAGTGAGTCAAAAAGGTACGCAGGTGATATCATTTTGGGTTT 840

Qy        841 AAGAGTTTTTTGTGAAGTTATTTTTTAATGATTTTTATTATTATTTTATTATTTAAAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGAGTTTTTTGTGAAGTTATTTTTTAATGATTTTTATTATTATTTTATTATTTAAAAT 900

Qy        901 CCTTTTTATTTCAGGGGAAAGTGGCAATCAATTTTTAGGCAAAGGTTTATTGATAACAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCTTTTTATTTCAGGGGAAAGTGGCAATCAATTTTTAGGCAAAGGTTTATTGATAACAT 960

Qy        961 AAAACTATGAAAGAAATTAAGGTATTGTTTATAGTTTAGGGTATGAGATAGATTTTGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAACTATGAAAGAAATTAAGGTATTGTTTATAGTTTAGGGTATGAGATAGATTTTGA 1020

Qy       1021 GGTTTTATTTGAAGATATGAATGGCATGGAATATGTTTATTTCTATGGGTACTTTAAAGTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GGTTTTATTTGAAGATATGAATGGCATGGAATATGTTTATTTCTATGGGTACTTTAAAGTT 1080

Qy       1081 AAAGGGTGGGCTGATTTAATTTGGTCAAATCCTAACTATATTCCTAATATATCATCCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AAAGGGTGGGCTGATTTAATTTGGTCAAATCCTAACTATATTCCTAATATATCATCCAG 1140

Qy       1141 AATTATTAAAGACGATGTTCCAAATTATCCTCTTGCTTTAATTAAAATGAGATTTAAGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATTATTAAAGACGATGTTCCAAATTATCCTCTTGCTTTAATTAAAATGAGATTTAAGCC 1200

Qy       1201 TTTTAGAGTTTTCAAAGTTACACAGTTGAAAAGTTAAAAATTTCATTTTTTATGTTAAAGA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTTAGAGTTTTCAAAGTTACACAGTTGAAAAGTTAAAAATTTCATTTTTTATGTTAAAGA 1260

Qy       1261 TTTAAGAGTTCTTTATGATAAGCTAAGTGTTTCAATAGATTCTGATATTGACAGTGAGTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTTAAGAGTTCTTTATGATAAGCTAAGTGTTTCAATAGATTCTGATATTGACAGTGAGTC 1320

Qy       1321 TGTATTTAAAGTTTATGAGACTAGCGGAACTGAATTTTTTTTTAAATTAAAGTTATATCTA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGTATTTAAAGTTTATGAGACTAGCGGAACTGAATTTTTTTTTAAATTAAAGTTATATCTA 1380

Qy       1381 AACTTTTAAAAGAGTTTTAAAGCTTAGAGAAAAAATTTCTATCGTTGAAGGCTTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AACTTTTAAAAGAGTTTTAAAGCTTAGAGAAAAAATTTCTATCGTTGAAGGCTTTTTTCA 1440

Qy       1441 AAACTTTGTAGAAAAGATTGAGAGTGAAAAACCTGAAGAATCATCTTTGAAAAATTAGGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAACTTTGTAGAAAAGATTGAGAGTGAAAAACCTGAAGAATCATCTTTGAAAAATTAGGT 1500

Qy       1501 TTAAATTAATATGTAAAGCTACCTAAAAGCTTTGCTTT[illegible]ACATATTAAAATAA 1560
              ||||||||||||||||||||||||||||||||||||||        ||||||||||||||
Db       1501 TTAAATTAATATGTAAAGCTACCTAAAAGCTTTGCTTT        ACATATTAAAATAA 1552

Qy       1561 TAGGAAATAGTATATGGAAATATTAGATTTGGAAAATGAAGAGCTTTTAGGAGTTTTTTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1553 TAGGAAATAGTATATGGAAATATTAGATTTGGAAAATGAAGAGCTTTTAGGAGTTTTTTT 1612

Qy       1621 TGAAGAAGCTCAAAATCTTGTAGATATCCTTGAAGAGAATATT 1663
              |||||||||||||||||||||||||||||||||||||||||||
Db       1613 TGAAGAAGCTCAAAATCTTGTAGATATCCTTGAAGAGAATATT 1655


RESULT 2
AAX61843
ID   AAX61843 standard; DNA; 1011 BP.
XX
```

[illegible]

... diseases
... Lyme disease ...

... polypeptides
... antibodies to members of
... particularly for the ... Lyme disease in
... be used for preventing or attenuating an
... of the Borrelia ... The products can also
...

[illegible]

Qy 1389 AAAGAG[illegible]AAAA[illegible] 1448

Db 902 AAAGNGIIIIAAAGCIIAGAGAAAAAAI[illegible]

Qy 1449 TAGAAAAGAI[illegible]GAAAAA[illegible] 1498

Db 962 [illegible]

RESULT [illegible]
AAX61844
ID AAX61844 standard; DNA; [illegible] BP.
XX
AC AAX61844;
XX
DT [illegible] (first entry)
XX
DE B. burgdorferi antigenic protein [illegible]
XX
KW Antigenic protein; vaccine; Lyme disease; [illegible]
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-[illegible]
PR 2[illegible]-JUN-1997; 97US-[illegible]
PR 22-JUL-1997; 97US-[illegible]
PR 22-JUL-1997; 97US-[illegible]
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX

```
PI   Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   P-PSDB; AAY20147.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 1; Page 214; 275pp; English.
XX
CC   This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence 945 BP; 300 A; 128 C; 192 G; 319 T; 6 other;

  Query Match             55.6%;  Score 924.6;  DB 20;  Length 945;
  Best Local Similarity   98.4%;  Pred. No. 3.6e-144;
  Matches  930;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        551 GAGGGTTCTAAAAGGGCAGAGCTGGAGAATTAGTCTTAGATTTTGCCGAGCTTGCAAGA 610
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db          1 GAGGGTTCTAAAAGGGCAGAGCTGGAGAATTAGTTTTAGATTTTGCCGAGCTTGCAAGA 60

Qy        611 GATCTAAGTTCAACTAGACTTGATCTTACAAATTATCTTGATTATGTATATTGGGCGGT 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GATCTAAGTTCAACTAGACTTGATCTTACAAATTATCTTGATTATGTATATTGGGCGGT 120

Qy        671 TCTGGTATTGTTAAGCCGGAAGATATGGTTGTAGATCTTGGGATAAATAATTGGAGCGTT 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTGGTATTGTTAAGCCGGAAGATATGGTTGTAGATCTTGGGATAAATAATTGGAGCGTT 180

Qy        731 TTACTTACTCTTCTGCAAGGTTGCAGGCTTACGTTAAAAATTCAGTTGTTGCGCCGCT 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTACTTACTCTTCTGCAAGGTTGCAGGCTTACGTTAAAAATTCAGTTGTTGCGCCGCT 240

Qy        791 CTTGTTAAGACTGAGTCAAAAAGGTACGCAGGTGATACTATTTTGGGGTAAGAGTTTTG 850
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db        241 CTTGTTAAGACTGAGTCAAAAAGGTACGCAGGTGATACTATTTTAGGGTAAGAGTTTTG 300

Qy        851 TTTCAAGCTATTCTCAATCATCTGCTATGATTATGGCACCATTTAAAATTGCTTTTTAT 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCAAGCTATTCTCAATCATCTGCTATGATTATGGCACCATTTAAAATTGCTTTTTAT 360

Qy        911 TCAGGGGAAACTGGCAATCAATTTTTAGGCAAAGGTCTTATTGATAACATTAAAACCATG 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCAGGGGAAACTGGCAATCAATTTTTAGGCAAAGGTCTTATTGATAACATTAAAACCATG 420

Qy        971 AAAGAAATTAAGGTATCTGTTTATAGTTTAGGGTATGAGATAGATCTTGAGGTTTTATTT 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAGAAATTAAGGTATCTGTTTATAGTTTAGGGTATGAGATAGATCTTGAGGTTTTATTT 480

Qy       1031 GAAGATATGAATGGCATGGAATATGCTTATTCTATGGGTACTTTAAAGTTTAAAGGGTGC 1090
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db        481 GAAGATATGAATGNCATGGAATATGCTTNNTCTATGGGTACTTTAAAGTTTAAAGGGTGC 540

Qy       1091 GCTGATTTAATTTGGTCAAATCCTAACTATATTCCTAATATATCATCCAGAATTATTAAA 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCTGATTTAATTTGGTCAAATCCTAACTATATTCCTAATATATCATCCAGAATTATTAAA 600

Qy       1151 GACGATGTTCCAAATTATCCTCTTGCTTCAAGTAAAATGAGATTTAAGGCTTTTAGAGTT 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GACGATGTTCCAAATTATCCTCTTGCTTCAAGTAAAATGAGATTTAAGGCTTTTAGAGTT 660

Qy       1211 TCAAAGTCACACAGTTCAAAAGTTAAAAATTTCATCTTTTATGTTAAAGATTTAAGAGTT 1270
              |||||||||||||||||||||     ||||||||||||||||||||||||||||||||||
Db        661 TCAAAGTCACACAGTTCAAAAGAGCAAAATTTCATCTTTTATGTTAAAGATTTAAGAGTT 720
```

↓

```
Qy       1271 CTTTATGATAAGTTGA[illegible]AATAGATT[illegible]TTTATTTAAA 13[illegible]
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTTTATGATAAGTTGAGGTTTCAATAGATTCGATATTGAGAGAGCTTTATTTAAA 7[illegible]

Qy       1331 CTTTATGAGACTAGGGAACTGAACCCTTCTAAATTAAAGGACAGAAA[illegible]AAA 13[illegible]
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTTTATGAGACTAGGGAACTGAACCCTTCTAAATTAAAGGACACNAANTTTAAA 84[illegible]

Qy       1391 AGAGTTTTAAAGCTTAGAGAAAAAATTTCTATG[illegible]GAA[illegible]AAAA[illegible]A 14[illegible]
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AGNGTTTTAAAGCTTAGAGAAAAAATTTCTATG[illegible]GAA[illegible]AAAA[illegible]A 9[illegible]

Qy       1451 CAAAAGATTGAGAGTGAAAAACCTGAAGAATCATCTCCGAAAAAT 1495
              |||||||||||||||||||||||||||||||||||||||||||||
Db        901 CAAAAGATTGAGAGTGAAAAACCTGAAGAATCATCTCCGAAAAAT 945


RESULT 4
AAX61740
ID   AAX61740 standard; DNA; 423 BP.
XX
AC   AAX61740;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein coding sequence, t2185nt.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057486.
PR   20-JUN-1997;   97US-0050350.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   P-PSDB; AAY20043.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 1; Page 171; 275pp; English.
XX
CC   This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence 423 BP; 138 A; 31 C; 71 G; 183 T; 0 other;

  Query Match             22.4%;  Score 372.8;  DB 20;  Length 423;
  Best Local Similarity   97.3%;  Pred. No. 1e-55;
  Matches  393;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 ATGATAATCTTTTTCAAAAAAGGTTTTTTATTTTCATTCTAGCAAGGGATTTGTTGCTA 60
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGATAATCTTTTTCAAAAAAGGTTTTTTATTTTCATTCTAGCAAGGGATTTGTTGCTA 79
```

diseases

Lyme disease in

attenuating an

The products can also

RESULT 6

XX

XX

(first entry)

XX

sheath protein

XX

KW

KW

XX

Treponema hyodysenteriae

XX

FH Key Location/Qualifiers

FT CDS

FT /*tag=

FT

FT /*tag=

FT sig_peptide

FT /*tag=

FT mat_peptide

FT /*tag=

FT -35_signal

FT /*tag=

FT -10_signal

FT /*tag=

FT RBS

FT /*tag=

XX

PN

XX

PD MAR

XX

PF SEP

XX

PR SEP

```
PR   24-JUL-1992;   92EP-0202273
XX
PA   (DUPH ) DUPHAR INT RES BV.
XX
PI   Koopman MBH,  Kusters JG;
XX
DR   WPI; 1993-102665/13.
DR   P-PSDB; AAR32279.
XX
PT   Vaccine to protect pigs against swine dysentery - comprises
PT   Treponema hyodysenteriae endo flagellum sheath protein, applied
PT   orally or intranasally
XX
PS   Claim 2; Page 21-22; 34pp; English.
XX
CC   This sequence encodes the endoflagellum sheath protein of T.
CC   hyodysenteriae.  The endoflagellum consists of at least four
CC   proteins, this protien forms the sheath of the flagellum and three
CC   proteins, of molecular weights 37, 34 and 32 kD, make up its core.
CC   Antibodies raised against the sheath protein have been shown to be
CC   bactericidal for T. hyodysenteriae.  The 43 kD sheath protein can
CC   be used in the production of a vaccine against infections such as
CC   swine dysentery
XX
SQ   Sequence 1140 BP; 390 A; 151 C; 234 G; 365 T; 0 other;

  Query Match             6.4%;  Score 107;  DB 14;  Length 1140;
  Best Local Similarity   49.3%;  Pred. No. 4.3e-09;
  Matches  368;  Conservative    ;  Mismatches  370;  Indels    ;  Gaps    ;

Qy        581 TTAGTCTTAGATTTTGCCGAGCTTGCAAGAGATCAAGTTCAACTAGACTTGATCTTACA 640
Db        136 TTTATCGCTGCATCTGCTTACGGTTTAACAAACTCAACTTTGATTGATTTTGCTTTAACA 195

Qy        641 AATTATTTTGATTACGTATATTC...........TAAGCCGGAAGATATGGTT 700
Db        196 GGTAATG.........GAGAAGGTGATA AAATGAA............... 255

Qy        701 GTAGATCTTGGGATAAATAATTGGAGGGTTTTACTTACTCCTTCTGCAAGGTTGCAGGCT 760
Db        256 GAAAATCTTTATAATGATAATT....................AATCT......... 315

Qy        761 TACTTTAAAAAT.................TAAAGAGTGA..AAAAAGGTA...A 820
Db        316 AATTGCAGAAATTCTTATTTACTAACGTAGACAGTAAAGGTAACAAGGTGCTTGGGAA 375

Qy        821 GGTGATACTATTTTGGGGGTAAGAGTTTTGTTTCCAAGCTATTCTC    AATCATCTGCT 877
Db        376 GCAGGTAAAGTTCTTGGTGTAAGAGTACATTTCCCATTAGCAGCTGGAACAGTTATGCT 435

Qy        878 ATGATTATGCCACCATTTAAAATCCTTTTTATTCAGGGAAAGTGGCAATCAATTTTT 936
Db        436 TTAGTAAAAACAGTATATGAACTTGAAATGTATGGCGGTGCTGATGGTACTAAATATACA 495

Qy        937   AGGCAAAGGTCTTATTGATAACATTAAAACCATGAAAGAAATTAAGGTATCTGTTTAT 994
Db        496 GAAGGTAAAGGGGTTATAGATAATGTTGGGGAAATAAATCTATTAGTTCTTGGGTTTAT 555

Qy        995 AGTTTAGGGTATGAGATAGATCTTGAGGTTTTATTTGAAGATATGAATGGCATGGAATAT 1054
Db        556 GGACGTAACTATTTAATTAGTTATTTCGTAAACTTACAAAATGAATTTGGTGAATTAAAA 615

Qy       1055 GTTTATTCTATGGGTATTTAAA...AAAGGGTGGGGTGATTTAATTGGTCAAATCCT 1114
Db        616 TCTTATCCAATGGGTACTGTTTACTTCAACGGTTGGAGACAAGTAAGATGGGAAAACAGA 675

Qy       1115 AACTATATTCCTAATATATCATCCAGAATTATTAAAGACGATGTTCCAAATTATCCTCTT 1174
Db        676 GAATACTTACCTAATGTTCCGCGACAGCGTATTAGTAAGAGAACCTCT    TTATCCTAGA 732

Qy       1175 TCTTCAAGTAAAATGAGATTTAATGTTTTTAGATTTTCAAAGTCACACAGTTCAAAAGTT 1234
Db        733 ATGATCCCTTCTGTTAAATTAGATTCTTAGGTTTCTATAGAACTAAAGATACTAAAGGC 792

Qy       1235 TAAAATTTTATCTTTTATGTTAAAGATTTAATAT..TTTATATAA..AA.......A 1294
Db        793 GGAGATTTCATCACTTACGTTAAAGATGTAACATTGAGTATGACTGAGTAGTTGTTGAT 852

Qy       1295 ATAGATTCTGATATTGACAGTGAGTCT 1321
Db        853 TTTGAAGAAGATATGGACGATGAAGCT 879


RESULT 7
ABL33071
ID   ABL33071 standard; DNA; 5294 BP.
XX
AC   ABL33071;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1044.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory, cancer, eye disease, arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-138609/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 1044; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5294 BP; 1264 A; 1153 C; 1439 G; 1438 T; 0 other;

  Query Match             4.7%;  Score 78.4;  DB 24;  Length 5294;
  Best Local Similarity   49.8%;  Pred. No. 0.0024;
  Matches  226;  Conservative    0;  Mismatches  226;  Indels    2;  Gaps    1;

Qy         56 TGCTAATTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATTTTTAGT 115
Db       3371 TGGTAATATATTAGTTATTCGGTTTGTTGGTT.AGATGGGAT...TGTTTTTATTTT 343

Qy        116 AAAAGATTTTCTTTTAGGAAGAATGAGGGTTTTCTTTTGGT...TATTAATGGATTT 175
```

[illegible]

CC [illegible] DNA [illegible]
CC cerebral-spinal fluid, tissue [illegible]
CC eyes, intestine, kidney, [illegible]
CC [illegible] slides, and [illegible]
CC sequences of the invention are useful [illegible]
CC diseases associated with [illegible]
CC cancer. ABK[illegible] ABK[illegible] represent [illegible] DNA
CC sequences of different genes associated with [illegible]
CC their complementary sequences.
CC Note: The sequence data for this patent [illegible] part of the printed
CC specification, but was obtained [illegible]
CC European Patent [illegible]
XX
SQ Sequence [illegible] BP; [illegible]

Query Match [illegible]; Score [illegible]; DB [illegible]; Length [illegible];
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

[illegible]

Qy 476 AAAAGAAA[illegible]AAAA[illegible]

AB[illegible]
ID AB[illegible] standard; DNA; [illegible]
XX
AC AB[illegible];
XX
DT 2[illegible] MAR [illegible] (first entry)
XX
DE Human immune system [illegible] N[illegible]
XX
KW Human; immune system disease; [illegible]
KW anti[illegible]
KW neuroprotective; anti HIV; anti[illegible]
KW antirheumatic; antiarthritic; anti[illegible]
KW antiinflammatory; [illegible] eye disease; [illegible]
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; [illegible]
KW [illegible].
XX
OS Homo sapiens.
XX

```
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation
XX
PS   Claim 1; SEQ ID NO 1369; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5641 BP; 1442 A; 34 C; 930 G; 3235 T; 0 other;

  Query Match             4.7%;  Score 78;  DB 24;  Length 5641;
  Best Local Similarity   47.7%;  Pred. No. 0.38828;
  Matches  229;  Conservative    0;  Mismatches  250;  Indels    0;  Gaps    0;

Qy          2 TGATAATCTTTTTCAAAAAAGGTTTTTTATTTTCATTCTAGCAAGGGATTTGTTGCTAA 61
Db       3950 TTAGAATTTTTAATATTTGATGGTAAGTATTATTTTGTTATTTTTTATTTAATGTTTT 4009

Qy         62 TTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATTTTTAGTAAAAGA 121
Db       4010 TTTAGTTTATTTATATGTTTATATTTAGATGAAATTTAGATTAGTTTTATTTTTTTAAAA 4069

Qy        122 TTTTCTTTTAGGAAGCAATGAGGGTTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGATT 181
Db       4070 TTTTGTTATAATTTTGATATTTAGTAAATTTTTAATTAATTTAAAAGAGTATATAGTTTG 4129

Qy        182 TTTATATTCTTTTTTAAAGAATGGAGTTTATTATGTAAATCTTTCAAGAGAATTTTATGA 241
Db       4130 GTGATATTTTTTTTTATATTAGAAATTTTGATATGTTTATTAATTTATTTAATTTTTATT 4189

Qy        242 TTCTTTTAATAATGGTGATTATAATGAATCTTTTGATGTTAAGGTCAATCTTTTTGCTAT 301
Db       4190 TTATTTTTTATTGGAGAATTTGAAGGTTTCGTTTGATTTTTGTTAATATTTGAGTGTTAT 4249

Qy        302 GTCTTTAATAAAAACAATGGGTTTAACTATCTGGTAAGATAAAAAAGATTATTATTCT 361
Db       4250 GATTTATTTGTTGATAGTGGGATGAGATAGTTTTTTATTATAATATTTTTTTTGATTTT 4309

Qy        362 TGTTGAAGGTGTATCTTAAAGGAGCAAAGTTGATAAATTAAATTTTACTAATAAAAATA 421
Db       4310 CGTTAAATATAATAATAATAGGTAATATTTTTGAGTGTTAATTGTTTATTAATTTTTTGT 4369

Qy        422 ATTAAAAAAGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTACATGAAAA 479
Db       4370 GTTAAGTATTTTATTATTTAATTTTTTTTTATTTTATAATTTATGATTCGTATGAAGA 4427


RESULT 14
AAT72882/c
ID   AAT72882 standard; cDNA; 19124 BP.
XX
AC   AAT72882;
XX
DT   12-SEP-1997  (first entry)
XX
DE   Plasmodium var-7 gene.
XX
KW   DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW   Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW   DABP; merozoite; malaria; var 1; var 2; var 3; var 7; immune response;
KW   Plasmodium; ss.
XX
OS   Plasmodium vivax.
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   exon            7377..15139
FT                   /*tag=  a
FT                   /number= 1
FT   intron          15140..16205
FT                   /*tag=  b
FT                   /number= 1
FT   exon            16206..17912
FT                   /*tag=  c
FT                   /number= 2
FT                   /note= "no stop codon given"
XX
PN   WO9641766-A2.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;  96WO-US09506.
XX
PR   07-JUN-1995;  95US-0487929.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Chitnis C,  Miller LH,  Peterson DS,  Su XZ,  ...
PI   Wellems TE;
XX
DR   WPI; 1997-052237/05.
DR   P-PSDB; AAW22475.
XX
PT   New malaria vaccines - contains cysteine rich DBL family protein
PT   binding domains homologous domains of the Duffy and sialic acid
PT   binding proteins
XX
PS   Claim 4; Page 56-61; 96pp; English.
XX
CC   This sequence represents the var 7 gene of Plasmodium. Var 7 belongs to
CC   the Duffy binding like (DBL) family of genes which have homology to the
CC   Duffy antigen binding protein (DABP) and sialic acid binding protein
CC   (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
CC   var family of genes modulate cytoadherence and antigenic variation of
CC   Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC   protein (DABP) are soluble proteins that appear in the culture
CC   supernatant after infected erythrocytes release merozoites. DABP and
CC   SABP mediate the binding of merozoites and schizonts to the erythrocyte
CC   surface. These proteins are necessary for erythrocyte invasion by the
CC   parasite. This sequence can be used in the compositions of the
CC   invention. The compositions are for the treatment and prevention of
CC   malaria, and comprise either a nucleotide sequence or encoded polypeptide
CC   of the var 1, var 2, var 3 or var 7 genes of the DBL gene family, a
CC   family of genes having homology with conserved regions of DABP and SABP.
CC   The compositions are used for the treatment and prevention of malaria.
CC   They are also used in the preparation of vaccines for inducing a
CC   protective immune response in a mammal to Plasmodium merozoites
CC   (especially Plasmodium falciparum or Plasmodium vivax).
XX
SQ   Sequence 19124 BP; 7824 A; 2190 C; 2786 G; 6324 T; 0 other;

  Query Match             4.7%;  Score 78;  DB 24;  Length 19124;
  Best Local Similarity   48.1%;  Pred. No. 0.389;
  Matches  222;  Conservative    0;  Mismatches  240;  Indels    0;  Gaps    0;
```

[illegible]

```
CC   Identical to the Duffy Antigen Binding Protein (DABP), and ... Antigen
CC   Binding Protein (SABP), which are soluble proteins that appear in the
CC   culture supernatant after erythrocytes infected with malaria release
CC   merozoites. Immunochemical studies indicate that DABP and SABP are the
CC   respective ligands for Plasmodium vivax and P. falciparum for the Duffy
CC   and sialic acid receptors on erythrocytes. The ... polypeptides can be
CC   used to vaccinate against malaria, especially caused by P. falciparum.
CC   Immunization with the polypeptide provides effective protection against
CC   malaria. The present sequence represents the ... polypeptide encoding
CC   DNA.
XX
SQ   Sequence [illegible]

  Query Match             4.[illegible]%;  Score [illegible]
  Best Local Similarity   4[illegible]%;
  Matches  222;  Conservative  [illegible];  Mismatches  [illegible]

Qy          1 ATGATAAT[illegible]AAAAAA[illegible]
Db      15842 ATTAT[illegible]TTTTTATTTAAATAAATT[illegible]AATA [illegible]

[illegible]

Qy        421 AATTAAAAA[illegible] 4[illegible]
Db      15412 AATATATATATAAAATATATATATAAAA[illegible]

RESULT 12
ABL[illegible]
ID   ABL[illegible] standard; DNA; [illegible] BP.
XX
AC   ABL[illegible];
XX
DT   26-MAR-[illegible]  (first entry)
XX
DE   Human metastasis associated gene [illegible]
XX
KW   Metastasis associated gene; [illegible]
KW   cytosine methylation; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200[illegible]-A1.
XX
PD   [illegible]-2002.
XX
PF   [illegible]-APR-2001; 2001WO-[illegible]
XX
PR   [illegible]-APR-2000; [illegible]
PR   07-APR-2000; 2000DE-[illegible]
```

```
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes,
PT   useful for diagnosis of cancers by analysis of cytosine methylation,
PT   also for treatment
XX
PS   Claim 1; SEQ ID No 127; 249pp; Sequence Listing; English.
XX
CC   The present invention provides a number of human metastasis associated
CC   genes which are modified by cytosine methylation. The sequences can be
CC   used in the diagnosis and treatment of cancer. The present sequence is
CC   one of the genes of the invention.
XX
SQ   Sequence 17491 BP; 4117 A; 284 C; 4472 G; 8618 T; 0 other;

  Query Match              4.7%;  Score 77.6;  DB 24;  Length 17491;
  Best Local Similarity   49.2%;  Pred. No. 0.00034;
  Matches  237;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy         50 ATTTGTTGCTAATTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCT 109
Db        155 ATTTAAATATGTTGTATTTTTTTAGAGATTTTTAGAATTTGGGAAGGGGAAATTATTTT 214

Qy        110 TTTAGTAAAAGATTTTCTTTTAGGAAGCAATGAGGGGTTTCTTTTGGGTTTTTATTAAG 169
Db        215 TTTAATTTTAGGAGGTTAAGATAGTTTTATAGAAGAATTTTTATTAATTTTTTAAACGAG 274

Qy        170 TGATTCAAGATTTTTATATTCTTTTTAAAGAATGGAGTTTATTATGTAAATCTTTCAAG 229
Db        275 TGGTTTTAGTTTTTGTTTTTGTTTTTTGTTAGTGTATTTTGTAAATGATTATTTTAAG 334

Qy        230 ATAATTTTATTATTCTTTTAATAAT...AAAC...ATAATAA...AA 289
Db        335 ATTAGGTATTTTTTGATATAATTTTTAGTTGATAGTTTAATATTTTTTATTAAAAGTTTA 394

Qy        290 TCTTTTGCTATGTTTTTAATAAAAACAATGCGCTTTAACTATCCTGGTAAGATAAAAAA 349
Db        395 AATAGTTGTAAAGGTTGTAATTTTTTGGTATATTGTAAATTTATATACGTAAAAGTAAAAT 454

Qy        350 GATTATTATTCTTGTTGAAGGGTGTATCTTAAAGGAGCAAAGTTGATAAATTAAATTTTA 409
Db        455 AATTATATTATTTTTTTAAATGTATTTAGTTAAATTAAATGTTTTAGTAATTCGATATTT 514

Qy        410 CTAATAA       AAATAATTAAAAAACGAAAATTTTATAAAAGATTTATATATAAGGAG 463
Db        515 ATAATTATTTATTTAAAAAAATATATGAATAAGTTTTTTTTTAAGAGTTAGATATTTTGTA 574

Qy        464 TTGTTTTACATGAAAAGGAAAGTAAAAGTATTTTATTTTTTTATTATCCACTGTTCTT 523
Db        575 TTATTTTTTTTTTATTTGAAGTAAATTTTTATTTTATGTTTTATTATAAAATTTTATT 634

Qy        524 TT 525
Db        635 AT 636


RESULT 14
ABL32323
ID   ABL32323 standard; DNA; 6418 BP.
XX
AC   ABL32323;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID No: 296.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200028-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-139909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation
XX
PS   Claim 1; SEQ ID No 296; 2pp; Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 6418 BP; 2072 A; 47 C; 1081 G; 3218 T; 0 other;

  Query Match              4.6%;  Score 75.8;  DB 24;  Length 6418;
  Best Local Similarity   50.5%;  Pred. No. 0.00049;
  Matches  215;  Conservative    0;  Mismatches  207;  Indels    4;  Gaps    1;

Qy        101 TTTAGATTTTTTATTAAAAGATTTTTTTTTATTAATTAATTAAGTGGTTTTCTTTTGTTTT 160
Db       2429 TTTTAATATGTTTTTTATTTATTTTATATTTTTTATATGTTTTTAGTTTTTGTTAGATGG 2488

Qy        161 TTTATTAAGTGATTCAAGATTTTTATATTCTTTTTAAAGAATGGAGTTTATTATGTAAA 220
Db       2489 ATTTTTATATTATTTGTTATTTTTAAAAATTTTTGAAAATAGAGTTAATAATTTTGATTT 2548

Qy        221 TCTTTTAAGAGAATTTTATGATTTTTTAATAATTTTATTATAATGAATTTTTTGATTT 280
Db       2549 TTATTTATTTAATTTTTTTATATTTTTTTTTTAAGATTTTATATAATTTTTTTTAAATTAT 2608

Qy        281 TAAGGTTAATTTTTTTGTTATTTTTAATAAAAA AA...AA TA...AA 340
Db       2609 ATTTATATATATTTTTTTTATAAATTAAAAATATATTTT     TTTTATAAAATATTTTTA 2664

Qy        341 GATAAAAAAGATTATTATTTTTTTTTAAATTTTTATTTTAAATTATTTAATTTTAGTAAAT 400
Db       2665 ATTAAATAAGAATTTTTATATTTTTTAAATATTTTATTTAAGAA...ATTTAT 2724

Qy        401 TAAATTTTATTAATAAAAATAATTAAAAAACGAAAATTTTATAAAAGATTTATATATAAG 460
Db       2725 GAATTTTTTAAAAATTTTATTATTGTTAGAGATAATTTTTAAAAAGATTATAATTTATAAT 2784

Qy        461 GAGTTGTTTTACATGAAAAGGAAAGTAAAAGTATTTTATTTTTTTTATTATCCACTGTT 520
```

[illegible] AAIAIAIAI [illegible] AIIIAAIIIIII 2844

[illegible]

[illegible] antiinflammat [illegible] antiasthmatic [illegible] cytostatic [illegible]

[illegible]

[illegible] immunodeficiency

[illegible]

[illegible]

Sq Sequence 64[illegible] BP; [illegible]

Query Match [illegible]
Best Local Similarity [illegible]
Matches [illegible] Conservative [illegible] Mismatches [illegible] Gaps [illegible]

```
Qy [illegible]
Db 24[illegible] [illegible]
Qy 161 [illegible]
Db 248[illegible] [illegible]
Qy 221 [illegible]
Db 254[illegible] [illegible]
Qy 281 [illegible]
Db 2609 [illegible]
Qy 341 [illegible]
Db [illegible]
Qy 4[illegible] [illegible]
Db 27[illegible] [illegible]
Qy 4[illegible] [illegible]
Db 278[illegible] [illegible]
Qy [illegible]
Db 2845 TTTATT 28[illegible]
```

RESULT 1[illegible]
ABL33278
ID ABL33278 standard; DNA; [illegible]
XX
AC ABL33278;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene [illegible]
XX
KW Human; immune system disease; [illegible]
KW antiarteriosclerotic; antianaemic; cytostatic; [illegible]
KW neuroprotective; anti-HIV; [illegible]
KW antirheumatic; [illegible]
KW antiinflammatory; cancer; [illegible]
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW [illegible] rheumatoid arthritis; [illegible]
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2002[illegible]28-A2.
XX
PD [illegible]-JAN-200[illegible].
XX
PF 02-JUL-2001; 2001WO-EP0[illegible]
XX
PR [illegible]-JUN-2000; [illegible]
PR [illegible]-SEP-2000; [illegible]
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI [illegible] A, [illegible] K, Berlin K;
XX

```
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation
XX
PS   Claim 1; SEQ ID NO 1251; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 7255 BP; 2118 A; 40 C; 1359 G; 3738 T; 0 other;

  Query Match             4.6%;  Score 76.6;  DB 24;  Length 7255;
  Best Local Similarity   48.3%;  Pred. No. 0.00048;
  Matches  274;  Conservative    0;  Mismatches  289;  Indels    4;  Gaps    2;

Qy          3 GATAATCTTTTTCAAAAAAGGTTTTTATTTTCATTCTAGCAAGGGATTTGTTGCTAAT 62
Db       4388 GTTAATTATAATTTTTTAGTAAATGTTTATTATTATTATTATTAGTATTGTTTTTTTAT 4447

Qy         63 TTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAGAT 122
Db       4448 TTTTTTTATTGAAATTTTAATTTATAGATTATTGATTTTTTAATATGTTGATATTAAATT 4507

Qy        123 TTTCTTTTAGGAAGCAATGAGGGTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGATTT 182
Db       4508 ATATTTTTTAGAAGTAGTTTTAGTTTTTATATAAAGTTTGTTTGTAATAATATAGTTAT 4567

Qy        183 TTATATTCTTTTT    TAAAGAATTGAGTTTATTATGTAAATCTTTCAAGAGAATTTTAT 239
Db       4568 TTATTTTTTTATATATAATGAGGGAGGGTAAGTATAGAGATTTTTTAGTTGGGTTTTTT 4627

Qy        240 GATTCTTTTAATAATGGTGATTATAATGAATCTTTTGATGTTAAGGTCAATCTTTTTGCT 299
Db       4628 GGTTTTAAGATATTTATTATAAATTTTAAAATATTTTTTTGTATAAGTTAATTTAATTTT 4687

Qy        300 ATTTCTTTTAATAAAAACAATGTGTTTTAATTATTTTTAAGATAAAAAAGATTATTATT 359
Db       4688 AATATTTTATATAGTTAAAAAATGTATTTATATAATTTTAGGATTAGTTTT TTGTTTTT 4746

Qy        360 CTTGTTGAAGGGTTATCTTAAAGGAGGAAAGTTGATAAATTAAATTTTACTAATAAAAA 419
Db       4747 TTTTTTTAATTAATTTTGTAAAAGGAAATAATTAAAAATATTTAAAAGTAAAGAGTAGAAA 4806

Qy        420 TAATTAAAAAACGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTACATGAAAA 479
Db       4807 GTTGGAGTTTTGAAAGATTGTTGAAAGATTTTTTTATTTGGGGTATTAAATTATATTAT 4866

Qy        480 GGAAAGCTAAAAGTATTTTATTTTTTTTATTATCCACTGTTCTTTTTGCTCAAGAGACTT 539
Db       4867 AAAATTGTAAAATTGTTAGAGTGTAATTAATTTATGTTGTTTATATTTTTGATGTAGTA 4926

Qy        540 ATGGATTAGCAGAGGGTTCTAAAAGGG 566
Db       4927 ATTGTTTAAGATATTGGTTTTAAAGAGT 4953

Search completed: November 4, 2002, 20:49:49
Job time : 484.738 secs
```

[illegible]

SUMMARIES

	Description
[illegible]	Sequence 13, Appl
[illegible]	Sequence 36, Appl
[illegible]	Sequence 288, App
[illegible]	Sequence 286, App
[illegible]	Sequence 186, App
[illegible]	Sequence 1147, Ap
[illegible]	Sequence 186, App
[illegible]	Sequence 1147, Ap
[illegible]	Sequence 1, Appli
[illegible]	Sequence 1, Appli
[illegible]	Sequence 1, Appli
[illegible]	Sequence 1, Appli
[illegible]	Sequence 13, Appl
[illegible]	Sequence 36, Appl
[illegible]	Sequence 27, Appl
[illegible]	Sequence 1, Appli
[illegible]	Sequence 1, Appli
[illegible]	Sequence 8, Appli
[illegible]	Sequence 8, Appli
[illegible]	Sequence 8, Appli
[illegible]	Sequence 780, App
[illegible]	Sequence 15, Appl
[illegible]	Sequence 11, Appl
[illegible]	Sequence 15, Appl
[illegible]	Sequence 13, Appl
[illegible]	Sequence [illegible], App
[illegible]	Sequence [illegible], Appl

[illegible]

ALIGNMENTS

RESULT 1
US-08-487-826B-1/[illegible]
; Sequence 1, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: [illegible] Kim L.
; APPLICANT: [illegible]
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: [illegible]
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 4[illegible]
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens [illegible]
; STREET: 620 Newport Center Drive [illegible]
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: [illegible]
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.[illegible]
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: [illegible]
; CLASSIFICATION: 4[illegible]
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: [illegible]
; REFERENCE/DOCKET NUMBER: NIH[illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: [illegible]
; TELEFAX: [illegible]
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: [illegible] base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

[illegible] Match [illegible]

```
  Best Local Similarity   48.1%;  Pred. No. 3.7e-07;
  Matches  222;  Conservative    0;  Mismatches  240;  Indels    0;  Gaps    0;

Qy          1 atgatgaatcttttttcaaaaaaaggtttttattttcattctagcaaggaatttgttgcta 60
Db      15832 ATTATTTTTTTTATTTAAATAAATTTTTTTTATTTTATGTATATATTTTTTTTTTAACA 15773

Qy         61 atttaaagatatttaagagatgaacaaatttgaaagataatttagatcttttagtaaaag 120
Db      15772 TTTTTTTAATTTTTTTTTATTTTATGATATATATTTTTATTTTAATATATTTTTTTTCTTT 15713

Qy        121 attttctttttaaaaaacaatgaaggatttttctttttagattttttattaaatgatttcaagat 180
Db      15712 TTTTTTGTTTTATGATATATATTTTTTTTTTTTTTTTTAATGTTTTTTTTTTCTTCTTTTG 15653

Qy        181 ttttatattctttttttaaaaaatgaagtttattatgtaaatctttcaagaaatttttatg 240
Db      15652 TTTTTATTTTTTTTTATAATCATTTTTTTTTTTATATAAAATTTTTTTTTAATTTTTTTTTG 15593

Qy        241 attcttttaataatgatgattatatgaatctttttgatgttaaggtcaatctttttgcta 300
Db      15592 ATAATTTTTTCATTTTTTATTTATCAAAATTTATATTTTATTATAATTTTTATTATTT 15533

Qy        301 tgtctttaataaaaacaatgcgctttaactatcctagtaagataaaaaagattattattc 360
Db      15532 TTTAAAAAATTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTA 15473

Qy        361 ttgttgaaggatgtatcttaaaggagcaaagttgataaattaaattttactaataaaaat 420
Db      15472 TATTTCATTTTTCTTTTTTCATTTTAATAAATTTGTTTTTATATTTTCTTTTTAAATATA 15413

Qy        421 acttaaaaactaaattttatataaaatttatatataagga 462
Db      15412 AATACATATATATAAAATATATATATAAAACACATACAAGTA 15371

RESULT   [illegible]
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
;  GENERAL INFORMATION:
;     APPLICANT:  Dekleve, Genevieve
;     APPLICANT:  Awang, Gregor
;     TITLE OF INVENTION:  Recombinant DNA Molecules and Expression
;     TITLE OF INVENTION:  [illegible]
;     NUMBER OF SEQUENCES:  [illegible]
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  BERESKIN & PARR
;       STREET:  40 King Street West
;       CITY:  Toronto
;       STATE:  Ontario
;       COUNTRY:  Canada
;       ZIP:  M5H 3Y2
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/883,795A
;       FILING DATE:  27-JUN-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gravelle, Micheline
;       REGISTRATION NUMBER:  40,261
;       REFERENCE/DOCKET NUMBER:  3641-062
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 364-7311
;       TELEFAX:  (416) 361-1398
;  INFORMATION FOR SEQ ID NO:  36:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  665 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       CLONE:  Rh 32
US-08-883-795A-36

  Query Match             3.9%;  Score 63;  DB 2;  Length 665;
  Best Local Similarity   47.8%;  Pred. No. [illegible];
  Matches  249;  Conservative    0;  Mismatches  [illegible];  Indels    0;  Gaps    0;

Qy          6 aatctttttttcaaaaaaaggtttttattttcattctagcaaggaatttgttgctaatttta 65
Db         33 AATATTTATAATTAAAATATTTATAATTAATATTTATAATTTATATTTTTTTAATATA 92

Qy         66 agatatttaaagagatgaacaaatttgaaagataatttagatcttttagtaaaagatttt 125
Db         93 AAATATTTA  TAATTAAATATTTATAATTAAAATATTTATAATTAAATATTTATAAT 152

Qy        126 cttttaggaagcaatgaaggattttctttttggatttttattaaatgattcaagattttta 185
Db        151 TAAAAATATTTATAATTAAATATTTATAATTAAAATATTTATAATTAAATATTTATAAT 210

Qy        186 tattctttttttaaaaaatgaagtttattatgtaaatctttcaagaaatttttatgattc 245
Db        211 TAAAAATATTTATAATTAAATATTTATAATTAAAATATTTATAATTAAATATTTATAAT 270

Qy        246 ttttaataatgatgattatatgaatctttttgatgttaaggtcaatctttttgctatgtc 305
Db        271 TAAAATATTTATAATTA      AATATTTATAATTAAAATATTTATAATTATATATTTT 324

Qy        306 tttaataaaaacaatgcgctttaactatcctagtaagataaaaaagattattattcttgtt 365
Db        [illegible] TTTAATTAAAATTTTTATAATTTATATTTTTATAATTAAAATATTTATAATTATATTTT [illegible]

Qy        366 gaaggatgtatcttaaaggagcaaagttgataaattaaattttactaataaaaataatta 425
Db        [illegible] ATAATTAAAATTTTTATAATTTATATTTTTATAATTAAAATATTTATAATTATATTTT 444

Qy        426 aaaaacgaaaatttttatataaaagatttatatataaggaatttagtttacataaaaagaaa 485
Db        [illegible] ATAATTATATATTTTTAAAATATTTTAATTATATATTTATAATTAAATATTTTTAA [illegible]

Qy        486 ctaaaaatatttttattttttttattatccactgttcttttt 527
Db        505 TACATATTTTATAATTAAAGTATTTATAATTACATATTTT 540

RESULT   [illegible]
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;     APPLICANT:  Philippsen, Peter
;     APPLICANT:  Pohlmann, Rainer
;     APPLICANT:  Steiner, Sabine
;     APPLICANT:  Mohr, Christine
;     APPLICANT:  Wendland, Jurgen
;     APPLICANT:  Knechtle, Philipp
;     APPLICANT:  Rebischung, Corinne
;     TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;     TITLE OF INVENTION:  AND USES THEREOF
;     NUMBER OF SEQUENCES:  1152
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  No. 6239264artin Corporation
;       STREET:  3054 Cornwallis Road
;       CITY:  Research Triangle Park
;       STATE:  No. 6239264th Carolina
;       COUNTRY:  USA
;       ZIP:  27709
```

[illegible]

```
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6234264
;  GENERAL INFORMATION:
;    APPLICANT: Philippsen, Peter
;    APPLICANT: [illegible]
;    APPLICANT: Steiner, Sabine
;    APPLICANT: [illegible] Christine
;    APPLICANT: [illegible]
;    APPLICANT: [illegible], Philipp
;    APPLICANT: [illegible]
;    TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION: AND USES THEREOF
;    NUMBER OF SEQUENCES: [illegible]
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: [illegible] Corporation
;      STREET: [illegible] Cornwallis Road
;      CITY: Research Triangle Park
;      STATE: North Carolina
;      COUNTRY: USA
;      ZIP: 277[illegible]
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #[illegible]
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/998,416
;      FILING DATE: 24 DEC 1997
;      CLASSIFICATION: [illegible]
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: [illegible]
;      FILING DATE: [illegible] DEC 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME: [illegible]
;      REGISTRATION NUMBER: [illegible]
;      REFERENCE/DOCKET NUMBER: [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: [illegible]
;      TELEFAX: [illegible]
;  INFORMATION FOR SEQ ID NO: 288:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: [illegible] base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM: [illegible]
US-08-998-416-288

  Query Match            [illegible]
  Best Local Similarity  [illegible]%;  Pred. No. [illegible]
  Matches  214;  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible]

Qy   [illegible]
Db   650 ATAATATA[illegible] 5[illegible]

Qy   [illegible]
Db   596 TAATTAATTAAATAATGATAAAAATATTAATTAAATAATTAATTTAATTAAATAAATAAAT 5[illegible]

Qy   [illegible]
Db   536 AAATAAAATTAAATA[illegible]TAATTTATTAAATAATATAAAAATTAATAATAAGAAATTAAAT 4[illegible]

Qy   [illegible]
Db   478 TTAAAATTAATTTTAATAATAATTTTTATAAAAAGATTAAATAATATAAATTAAGATAAT 41[illegible]

Qy   [illegible]
```

```
Db     418 ATTTATAAAAATAGATATTATAATAAAAATAATAATTTATAATATTTAA     ATAAATT 363

Qy     402 aaattttactaataaaaataatttaaaaaacgaaaattttataaaagatttatatataagg 461

Db     362 AAATCTTTATAATAATAAAAATATATTATTTTAATAAACAATTAATAAATAATATATTAAT 303

Qy     462 agttggtttacatgaaaaggaaagctaaaagtattttattttttttatta 511

Db     302 TATTGATAATCTATTTAATAATTTATTAAAGAAAATAATAATATCTAATA 253


RESULT 5
US-08-998-416-186
; Sequence 186, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3018/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  [illegible] 1974
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  186:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  EAC274EF
US-08-998-416-186


  Query Match             3.7%;  Score 61.8;  DB 4;  Length 615;
  Best Local Similarity   47.8%;  Pred. No. [illegible];
  Matches  211;  Conservative    0;  Mismatches  227;  Indels    3;  Gaps    1;

Qy      75 agagatgaacaaaatttgaaaaataatttagatctttagtaaaagatttcttttagga 134

Db      10 ATATATAATTTAATTTATAAAGATTAATATAAACTTTTTTATTATAATATTTAAGTATTA 69

Qy     135 agaaatgaagagatttttcttttggqtttttatttaagtaat [illegible] 191

Db      70 AATTATTTAAACTATTATTATCATTATTTAATAAATTAATTATTTGATTATTAATACTTA 129

Qy     192 ttttaaagaatggagtttatttatgtaaatctttcaaaagaatttt [illegible] 251

Db     130 TTATATAATTATTATATAATTTATTTAATTTATTATTATAATATTTATTATAATTATTAA [illegible]

Qy     252 aatggtgatttataatgaatctttttgatgttaaagtcaatctttttgctatatcttaat [illegible]

Db     190 AATAATATTTAATATGAATACTATTTATTTCTATTTTCAAATTTTAAATTAGTTATTAAAA [illegible]

Qy     312 aaaacaatgcgctttaactatcctggtaagataaaaaaaattatttatttcttgttaaaaa 371

Db     2?  TATTTATTATATATTATTATTTTTTTTAATAAATTTTTATTTTTTTTTTTATAATTAATA [illegible]

Qy     372 tgtatcttaaaaggagcaaagttgataaattaaattttactaataaaaataattaaaaaac 431

Db     310 TATTATTTATTAATTTTTTATTAAAAATAATATATTTTATTATTATAAAGATTTAATTTAT [illegible]

Qy     432 caaaattttataaaagatttatatataaggagttggtttacatgaaaaggaaagctaaaa 491

Db     370 TTAAATATTGTAAATTATTATTTTTATTATAATATTTATTTTTATAAATATTATTTTTAT 42?

Qy     492 gtattttattttttttattat 512

Db     430 TTATATTATTTAACTTTTTAT 45?


RESULT 6
US-08-998-416-1137
; Sequence 1137, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3018/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1137:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
```

```
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #[illegible] Version #[illegible]
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/[illegible]
;       FILING DATE: 24-[illegible]
;       CLASSIFICATION: 435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: [illegible]
;       FILING DATE: [illegible]
;     ATTORNEY/AGENT INFORMATION:
;       NAME: [illegible]
;       REGISTRATION NUMBER: [illegible]
;       REFERENCE/DOCKET NUMBER: [illegible]
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: [illegible] 541 [illegible]
;       TELEFAX: [illegible] 541 [illegible]
;   INFORMATION FOR SEQ ID NO: [illegible]
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: [illegible] base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: [illegible]
;       TOPOLOGY: [illegible]
;     MOLECULE TYPE: DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM: [illegible]
US-08-998-416-186

  Query Match [illegible]; Score [illegible]; DB [illegible]; Length [illegible];
  Best Local Similarity [illegible]; Pred. No. [illegible];
  Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Db        256 AATA 25[illegible]

RESULT [illegible]
US-08-998-416-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. [illegible]
;  GENERAL INFORMATION:
;    APPLICANT: [illegible]
;    APPLICANT: [illegible], Rainer
;    APPLICANT: [illegible]
;    APPLICANT: [illegible]
;    APPLICANT: Weinand[illegible], [illegible]
```

```
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,116
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3016/96
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  [illegible]76
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  [illegible]
;  INFORMATION FOR SEQ ID NO:  1147:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAC1692BF
US-08-998-116-1147

  Query Match             [illegible]%;  Score 59.2;  DB 4;  Length 636;
  Best Local Similarity   48.3%;  Pred. No. 0.001;
  Matches  195;  Conservative    0;  Mismatches  208;  Indels    1;  Gaps    1;

Qy     55 ttgctaatttaagatatttaagagatgaacaaatttgaaagataatttagatcttttag 114
Db    407 TAGATATTATAATAAAAATAATAATTTACAATATTTAAATAAATAAAT[illegible]TTTATAATAA 348

Qy    115 taaaagattttcttttaggaagcaatgaggggttttcttttgggtttttattaagtgatt 174
Db    347 [illegible] 288

Qy    175 caagattttttatattcttttttaaaaaatggagtttatttatgtaaatctttcaaaagaat 234
Db    287 TAATAATTTATTAAAGAAAATAATAATATCTAATA-ATATTTTAATAACTAATTTAAAAT 229

Qy    235 tttatgattcttttaataatggtgattataatgaatcttttgatgttaaggtcaatcttt 294
Db    228 TTGAACATAGACTAAATAGTATTCATATTAAATATTATTTTTATAATTATATAAATATTA 169

Qy    295 ttgctatatgtctttaataaaaacaatgcgctttacctatcctggtaagataaaaagatta 354
Db    168 [illegible] 109

Qy    355 ttattcttgttgaagggtgtatcttaaaggagcaaagttgataaattaaatttactaat 414
Db    [illegible]

Qy    415 aaaaataattaaaaaacaaaaatttatttataaaagatttatatata 458
Db     46 AAAAAGTTTATATTAATTTTATAAATTAAATTATATATATTTA 3

RESULT 9
US-08-446-355A-1/c
; Sequence 1, Application US/08446355A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence of [illegible]
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,355A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29,009
;      REFERENCE/DOCKET NUMBER:  47-90
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic
US-08-446-355A-1

  Query Match             [illegible]%;  Score [illegible];  DB [illegible];  Length 3920;
  Best Local Similarity   44.4%;  Pred. No. [illegible];
  Matches  224;  Conservative    [illegible];  Mismatches  28[illegible];  Indels  [illegible];  Gaps  [illegible];

Qy      8 tcttttttcaaaaaaggttttttattttcattctagcaaggagatttattgctaatttaag 67
Db    [illegible]

Qy     68 atatttaagagatgaacaaaatttgaaaagataatttagatcttttatagaaaattttt[illegible] 127
Db    674 ATTTTTCATATGTAAATTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAA [illegible]

Qy    128 tttaggaagcaatgaggggtttttcttttgggtttttattaagtgattcaagattttatat [illegible]
Db    614 ATTAGTAGAATTACTATTTTATAAACTAAGAAAAAAAATAAATAAATGAATAAATTAATA [illegible]

Qy    188 ttcttttttaaaaaatggagtttatttatgtaaatctttcaaaagaattttatgattcttt [illegible]
Db    5[illegible]4 AATAAAATATATAAATAAAATTATATTAATAATAATATTTTATATTATATATTTTTTA 4[illegible]

Qy    248 taataatggtgattataatgaatcttttgatgttaaggtcaatcttttgctatatgtcttt [illegible]
Db    [illegible] AA 4[illegible]

Qy    308 aataaaaacaatgcgctttacctatcctggtaaagataaaaagattattattcttgttaa [illegible]
```

```
[illegible]

RESULT [illegible]
US-[illegible]
; Sequence [illegible], Application US/[illegible]
; Patent No. [illegible]
;  GENERAL INFORMATION:
;    APPLICANT: [illegible]
;    APPLICANT: [illegible]
;    APPLICANT: Hedstrom, Richard
;    APPLICANT: [illegible]
;    APPLICANT: [illegible]
;    TITLE OF INVENTION: [illegible]
;    TITLE OF INVENTION: [illegible]
;    NUMBER OF SEQUENCES: [illegible]
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: A. David Spevack
;      STREET: NMRI [illegible]
;      STREET: Medical Center
;      CITY: Bethesda
;      STATE: MD
;      COUNTRY: USA
;      ZIP: [illegible]
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM [illegible]
;      OPERATING SYSTEM: [illegible]/MS-DOS
;      SOFTWARE: PatentIn Release #1.[illegible]
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/[illegible]
;      FILING DATE: [illegible]
;      CLASSIFICATION: 424
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Spevack, A[illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: [illegible]
;      TELEFAX: [illegible]
;  INFORMATION FOR SEQ ID NO: [illegible]:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: [illegible] base pairs
;      TYPE: NUCLEIC ACID
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;    MOLECULE TYPE: DNA (genomic)
;    HYPOTHETICAL: N
;    ANTI-SENSE: N
;    ORIGINAL SOURCE:
;      ORGANISM: [illegible]
;      STRAIN: [illegible]
;      DEVELOPMENTAL STAGE: erythrocytic stage
;      TISSUE TYPE: Blood
;      CELL TYPE: erythrocytic stage
;    IMMEDIATE SOURCE:
;      LIBRARY: [illegible]
;      CLONE: [illegible]
;    FEATURE:
;      NAME/KEY: CDS
;      LOCATION: [illegible]
;      OTHER INFORMATION:
US-[illegible]

  Query Match [illegible]
  Best Local Similarity [illegible]
  Matches [illegible]
```

```
Qy          1 atgataatcttttttcaaaaaaggtttttttattttcattctagcaaggatttgttgcta 60
Db       3910 ATCATGAGCCAACTTTAGATATTCTTTTTTTAACATTCATCATTTTTTTTTATCACAC 3851

Qy         61 atttaagatatttaagagatgaacaaaatttgaaagataatttagatcttttagtaaaag 120
Db       3850 TTTTTAGTTTATAAAATTAAGACCATTATTTTTATGTTTTGTAAATTTTAAAAATAATCA 3791

Qy        121 atttttcttttaggaagcaatgagggtttttcttttgggttttttattaagtgattcaagat 180
Db       3790 CATTTTGTTTATTTTCTTTTTTATCGATAATATTGGTGATTTGTCTATTTTTTTAGGAAT 3731

Qy        181 ttttatattctttttttaaagaatggagtttattatgtaaatctttcaagagaattttatg 240
Db       3730 TCATTTTATTATGTATTATCACTTTTTTGTTTTATTCATAATTATTTTGAAAATAGTAAA 3671

Qy        241 atttcttttaataatggtgattataatgaatcttttgatgttaagtcaatcttttgcta 300
Db       3670 TACCGTGTAAATATACAAACCTAAAAATGTTATTAACTTTTAAGTTTTTTTTTTTTTTTT 3611

Qy        301 tatctttaataaaacaatgcgctttaactatcctgataagataaaaaagattattattc 360
Db       3610 TTTTTTTATATTAAGAATAATTGTAACCATTAAATATTGGAGTATAAATAAATATATAT 3551

Qy        361 ttgttgaaggtgtatct taaaaagcaaagttgataaattaaattttactaataaaa 419
Db       3550 ATTATAACGAGACAATTAGTTAAAAAAAAATAGTTAAAAAAAAATCGTTAAAAAAAAATA 3491

Qy        420 taattaaaaacgaaatttttataaaagatttatatataaggagttggtttacatgaaaa 479
Db       3490 TGAAAATAAATGGATATATAATTGAATGAATAACATAAAAAGATGACAATTTATCAAACT 3431

Qy        480 aggaaagctaaaagtattttat 500
Db       3430 GTTAATTTAAAATAACTTAAT 3410
```

RESULT 12
PCT US92 00018 1/c
; Sequence 1, Application PC/TUS9200018
; GENERAL INFORMATION:
;   APPLICANT: Hoffman, Stephen L.
;   APPLICANT: Charoenvit, Yupin
;   APPLICANT: Hedstrom, Richard
;   APPLICANT: Khusmith, Srisin
;   APPLICANT: Rogers IV, William O.
;   TITLE OF INVENTION: Protective malaria sporozoite surface protein
;   TITLE OF INVENTION: immunogen and gene encoding
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: A. David Spevack
;   STREET: NMRDC Building 1 1 12 National Naval
;   STREET: Medical Center
;   CITY: Bethesda
;   STATE: MD
;   COUNTRY: USA
;   ZIP: 20814 5044
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC DOS/MS DOS
;   SOFTWARE: PatentIn Release #1.24
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US92/00018
;   FILING DATE: 19920103
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Spevack, Avram D.
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 295 6750
;   TELEFAX: (301) 295 4033
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4673 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: N
;   ANTI SENSE: N
;   ORIGINAL SOURCE:
;   ORGANISM: Plasmodium yoelii
;   STRAIN: 17X(NL)
;   DEVELOPMENTAL STAGE: erythrocytic stage
;   TISSUE TYPE: Blood
;   CELL TYPE: erythrocytic stage
;   IMMEDIATE SOURCE:
;   LIBRARY: Py lambdagt11 2 7 Kb genomic expression
;   CLONE: Py10.1111
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 718..3195
;   OTHER INFORMATION:
PCT US92 00018 1

  Query Match   3.3%; Score 59.4; DB 5; Length 4673;
  Best Local Similarity   45.7%; Pred. No. 99;
  Matches 229; Conservative   0; Mismatches 271; Indels   0; Gaps   0;

```
Qy          1 atgataatcttttttcaaaaaaggtttttttattttcattctagcaaggatttgttgcta 60
Db       3910 ATCATGAGCCAACTTTAGATATTCTTTTTTTAACATTCATCATTTTTTTTTATCACAC 3851

Qy         61 atttaagatatttaagagatgaacaaaatttgaaagataatttagatcttttagtaaaag 120
Db       3850 TTTTTAGTTTATAAAATTAAGACCATTATTTTTATGTTTTGTAAATTTTAAAAATAATCA 3791

Qy        121 atttttcttttaggaagcaatgagggtttttcttttgggttttttattaagtgattcaagat 180
Db       3790 CATTTTGTTTATTTTCTTTTTTATCGATAATATTGGTGATTTGTCTATTTTTTTAGGAAT 3731

Qy        181 ttttatattctttttttaaagaatggagtttattatgtaaatctttcaagagaattttatg 240
Db       3730 TCATTTTATTATGTATTATCACTTTTTTGTTTTATTCATAATTATTTTGAAAATAGTAAA 3671

Qy        241 atttcttttaataatggtgattataatgaatcttttgatgttaagtcaatcttttgcta 300
Db       3670 TACCGTGTAAATATACAAACCTAAAAATGTTATTAACTTTTAAGTTTTTTTTTTTTTTTT 3611

Qy        301 tatctttaataaaacaatgcgctttaactatcctgataagataaaaaagattattattc 360
Db       3610 TTTTTTTATATTAAGAATAATTGTAACCATTAAATATTGGAGTATAAATAAATATATAT 3551

Qy        361 ttgttgaaggtgtatct taaaaagcaaagttgataaattaaattttactaataaaa 419
Db       3550 ATTATAACGAGACAATTAGTTAAAAAAAAATAGTTAAAAAAAAATCGTTAAAAAAAAATA 3491

Qy        420 taattaaaaacgaaatttttataaaagatttatatataaggagttggtttacatgaaaa 479
Db       3490 TGAAAATAAATGGATATATAATTGAATGAATAACATAAAAAGATGACAATTTATCAAACT 3431

Qy        480 aggaaagctaaaagtattttat 500
Db       3430 GTTAATTTAAAATAACTTAAT 3410
```

RESULT 13
US 08-487 826 13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
;   APPLICANT: Sim, Kim L.
;   APPLICANT: Chitnis, Chetan
;   APPLICANT: Miller, Louis H.

[illegible]

M VIVAX

AN [illegible] BINDING [illegible]

[illegible]

RESULT 14
US-09-004-395A [illegible]
; Sequence [illegible], Application US/ [illegible]A
; Patent No. [illegible]
; GENERAL INFORMATION:
; APPLICANT: [illegible]
; APPLICANT: [illegible]
; TITLE OF INVENTION: [illegible] DNA [illegible]
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: [illegible]
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: [illegible]
; STREET: 4[illegible] King Street West
; CITY: Toronto
; STATE: [illegible]
; COUNTRY: Canada
; ZIP: M[illegible]
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.[illegible]
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,[illegible]A
; FILING DATE: [illegible]
; CLASSIFICATION: [illegible]
; ATTORNEY/AGENT INFORMATION:
; NAME: [illegible]
; REGISTRATION NUMBER: [illegible]
; REFERENCE/DOCKET NUMBER: [illegible]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (41[illegible]) [illegible]
; TELEFAX: (41[illegible]) [illegible]
; INFORMATION FOR SEQ ID NO: [illegible]:
; SEQUENCE CHARACTERISTICS:
; LENGTH: [illegible] base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: [illegible]
US-09-004-395A [illegible]

Query Match [illegible]; Score [illegible]; DB [illegible]; Length [illegible];
Best Local Similarity [illegible]; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Qy [illegible]

Db 454 [illegible]

Qy [illegible]

Db [illegible]

Qy [illegible]

Db [illegible]

Qy [illegible]

Db 274 [illegible]

Qy [illegible]

Db 214 [illegible]

Qy [illegible]

```
Db    154 [illegible] 95

Qy    [illegible] 452

Db     94 TTTAATTATAAAATATTTAATTATAAATATTTTAATTATAAATATTTTAATTATAAAATA 35

Qy    453 tata 456

Db     34 TTTA 31
```

RESULT 15
US-08-817-926-27
; Sequence 27, Application US/08817926
; Patent No. 6091598
; GENERAL INFORMATION:
; APPLICANT: Komeda, Toshihiro
; APPLICANT: Sudo, Hisako
; APPLICANT: Tamai, Yukio
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Kato, No. 6001590no
; APPLICANT: Sakai, Yasuyoshi
; TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINII
; TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,926
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/02597
; FILING DATE: 12-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 234133/1995
; FILING DATE: 12-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 42538/1996
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 81356/6112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4818 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Candida boidinii
US-08-817-926-27

Query Match 3.2%; Score 54; DB 3; Length 4818;
Best Local Similarity 50.0%; Pred. No. 0.016;
Matches 162; Conservative 0; Mismatches 160; Indels [illegible]; Gaps [illegible];

```
Qy    173 [illegible]
Db   4376 [illegible] 44[illegible]

Qy    231 [illegible]
Db   4436 [illegible] 44[illegible]

Qy    291 [illegible]
Db   4495 [illegible]

Qy    351 [illegible] 41[illegible]
Db   4556 [illegible]

Qy    411 [illegible] 47[illegible]
Db   4616 [illegible]

Qy    473 [illegible] 494
Db   4675 [illegible] 46[illegible]
```

Search completed: March [illegible]
Job time: [illegible] sec

[illegible] Seconds

[illegible] updates/sec

Sequence 11234, A
Sequence [illegible], Appl
Sequence [illegible], Appl
Sequence 10265, A
Sequence [illegible], Appl
Sequence 4582, Ap
Sequence 4584, Ap
Sequence [illegible], Ap
Sequence 11234, A
Sequence 5785, Ap
Sequence [illegible], Ap
Sequence [illegible], Appl
Sequence 13489, A
Sequence 13781, A
Sequence 15614, A
Sequence 4584, Ap
Sequence [illegible], Appl
Sequence [illegible], Appl
Sequence 4637, Ap

ALIGNMENTS

RESULT 1
US-09-960-352-11234
; Sequence 11234, Application US/[illegible]
; Patent No. US2[illegible]A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley [illegible]
; APPLICANT: [illegible], Nozahiro
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Natappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: [illegible]
; CURRENT APPLICATION NUMBER: US/[illegible]
; CURRENT FILING DATE: [illegible]
; NUMBER OF SEQ ID NOS: [illegible]
; SEQ ID NO [illegible]11234
; LENGTH: 41[illegible]
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: [illegible]
US-09-960-352-11234

Query Match [illegible]; Score [illegible]; DB [illegible]; Length [illegible];
Best Local Similarity 49.4%; Pred. No. [illegible];
Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Qy 147 [illegible]

Db [illegible]

Qy [illegible]

Db 97 [illegible]

Qy 2[illegible] [illegible]

Db 1[illegible] [illegible]

Qy [illegible]27 [illegible]

Db [illegible]1[illegible] [illegible]

```
Qy    [illegible]
Db    277 TTTTTAAAATTTTTTAAAAAAATTATAATTTAAATAAAATAATAAATAAAAAATTTTAAAA 336

Qy    447 [illegible] 506
Db    337 AAAATTTTTTAAAAAAATATTTTTATAAAAATAATTTTTATATAAATAATTTTATTTTTT 396

Qy    507 TA 508
Db    397 TA 398

RESULT 2
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT: WATANABE, HIDEMI
;  APPLICANT: HATTORI, MASAHIRA
;  APPLICANT: SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE: 2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107165
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             3.7%;  Score 61;  DB 10;  Length 640681;
  Best Local Similarity   45.4%;  Pred. No. 0.046;
  Matches  225;  Conservative    0;  Mismatches  255;  Indels    [illegible];  Gaps    [illegible];

Qy     25 TTTTTTATTTTCATTCTAGGAAGGGATTTGTTGGTAATTTAAGATATTTAAGAGATGAAT 84
Db 486554 TATTTTATATTTTTTTAGGAGAAGATTTTTTTTTATTAGAAAAAAATCAAGATATGATA 486613

Qy     85 [illegible] 144
Db 486614 TTGAATTTTGATATAAAAAAGGATT[illegible]TAGAAAAAAATTATAATTGATGTAGAAAAAAAT 486673

Qy    145 [illegible] 204
Db 486674 [illegible]

Qy    205 GACTTTATTATGTAAATCTTTCAAGAGAATTTTATGATTCTTTTAATAATGGTGATTATA 264
Db [illegible]

Qy    265 [illegible] 324
Db 486794 AAAATGTTTTTCTTTATTGCATCCAGATATATTAATAATATTAAAATTCAATCATTTATCA 486853

Qy    325 [illegible] 384
Db 486854 [illegible]

Qy    385 [illegible] 444
Db 486914 [illegible] 486973

Qy    445 [illegible] 504
Db 486974 ATAAATATTGAGGAAAAAGCATTTTTTTTATTATGTAAATATTATGAAGGTAATACTTTA 487033

Qy    505 TTTAT 509
Db 487034 TTTAT 487038

RESULT 3
US-09-801-861-3
; Sequence 3, Application US/09801861
; Patent No. US20020119544A1
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: CL001098
;  CURRENT APPLICATION NUMBER: US/09/801,861
;  CURRENT FILING DATE: 2001-03-09
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: [illegible]
;   TYPE: DNA
;   ORGANISM: Human
US-09-801-861-3

  Query Match             3.7%;  Score 60.8;  DB [illegible];  Length [illegible];
  Best Local Similarity   48.1%;  Pred. No. [illegible];
  Matches  17[illegible];  Conservative    [illegible];  Mismatches  18[illegible];  Indels    [illegible];  Gaps    [illegible];

Qy      8 [illegible]
Db  31312 TATATATTTTATATATATTTATATTATATATTTTATATATATTTATATATTTATATAT 31371

Qy     68 [illegible]
Db  31372 TATATTTTATATATTTATATATTTATATATATTTTATATATTTATATATTTATATATT [illegible]

Qy    128 [illegible]
Db  31432 TTTATATATTTATATATTTATATATTTTATATATTTTATATATTTATATATATTTATATA [illegible]

Qy    188 [illegible] 24[illegible]
Db  31492 TATTTATATATATATTATATATATTTATATTTATATATATTTATATATTTATATATATTT [illegible]

Qy    248 [illegible]
Db  31552 ATATATTTATATATATTATATATTTTATATATTTATATATTATATATATTTATATATTT [illegible]

Qy    308 [illegible] 367
Db  31612 ATATATTTATATATTATATATATTTTTTTATATATATATATATTATTTTTTTTTTGA [illegible]

RESULT 4
US-09-960-352-10265
; Sequence 10265, Application US/09960352
; Patent No. [illegible]A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT: Tao, Nengbing
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND [illegible] WITH [illegible] AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: [illegible]
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE: 2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10265
;   LENGTH: 406
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 44 [illegible]
```

```
Qy     641 AATIATTTTATTATTTATATT 661

Db   [illegible] TAATTATTATTTTATTTTTT [illegible]


RESULT [illegible]
US-[illegible]-4582
; Sequence 4582, Application US/[illegible]
; Patent No. US2[illegible]A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley [illegible]
;  APPLICANT: [illegible]
;  APPLICANT: [illegible]
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND [illegible] ASSOCIATED WITH [illegible] AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: [illegible]
;  CURRENT APPLICATION NUMBER: US/[illegible]
;  CURRENT FILING DATE: 2[illegible]-24
;  NUMBER OF SEQ ID NOS: [illegible]
; SEQ ID NO 4582
;   LENGTH: [illegible]
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: [illegible]
US-[illegible]-4582

  Query Match             [illegible]
  Best Local Similarity   [illegible]
  Matches [illegible]  Conservative [illegible]  Mismatches [illegible]  Indels [illegible]  Gaps [illegible]

Qy     148 TTTTTTTGGTTTTATTAAGTGATTTAATATTTTATATTTTTTTTAAAAAATTGAT [illegible]

Db     [illegible] TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT [illegible]

Qy     208 TTTATTATTAAATTTTTAAGAAATTTTATTATTTTTTAATAATTTTATTATAATT [illegible]

Db     [illegible] TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT [illegible]

Qy     [illegible] AA[illegible]AAAAATAATTTTTTA [illegible]

Db     18[illegible] TTTTTATTTTTATTTATTTTATAATTTTTTATTTTTAAAAAATTATTTATAAAATTTT [illegible]

Qy     [illegible] ATTATTTTTTAATTAAAAAA[illegible]

Db     24[illegible] TTTTTTTTTTTAAATATATATTTTTAAAAAATTATTAAAATAATTATTATATTTTA [illegible]

Qy     [illegible] AAATT[illegible]AAAAT 44[illegible]

Db     [illegible] TTATTATATTATTTTATTATTATAAATTTAAATATTTTATAAATTTTTTTTTTATATT [illegible]

Qy     448 ATTTATATATAATTATTTTTTTATATTAAA [illegible]
```

```
Db        390 AATTTTTTAAATTAATTTTTTATTTTTTTAA 390

RESULT 7
US-09-960-352-4584/c
; Sequence 4584, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT: Jao, Nenabina
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(19298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE: 2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4584
;   LENGTH: 416
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 20 LIB3657-016-Q1-K1-E11
US-09-960-352-4584

  Query Match             3.3%;  Score 55.6;  DB 10;  Length 416;
  Best Local Similarity   47.8%;  Pred. No. 0.08;
  Matches  199;  Conservative    0;  Mismatches  209;  Indels    8;  Gaps    1;

Qy         90 TATTTTCATTCTAG[illegible] 89
Db        408 TTTTTTTTTTAATTAAAAATTTTTATTATAAAATTTTTAATTTATAAAATTACCCATA 349

Qy         90 TTGAAAGATAATTTAGATTTTTAGTAAAAGATTTTTTTTTAGGAAGTAATGAGGGTTT 149
Db        348 TTATTTTTAAATATTTATTTTTAATAAATCTTTTTTTATATCTTCTATAATTTAATAATT 289

Qy        150 TCTTTTGGGTTTTTATTAAGTGATTCAAGATTTTTATATTCTTTTTTAAAGAATGGAGTT 209
Db        288 TTAAAATTTTTAAAAATTTTTTTTAAAAATTTAAAAATTTTTTTTTTTAAATATTTTTT 229

Qy        210 TATTATGTAAATTTTTAAGAGAATTTTATGATTTTTAATAATGATGATTATAATGAA 269
Db        228 TTTTTATGAATTTTTTAATTTTTATATTTTAATTTTTAAAATCTTTTTTTTTTTTTTTT 169

Qy        270 TCTTTTGATTTTAATTTAATTTTTTTTTATGTTT[illegible]AATAAAAAGAATGGGTTTAAT 329
Db        168 TTTTTTTTTTTTAATTTTATTTTTTTT      TTTTATAAAAATTTATTTTTTTATT 117

Qy        330 TATTTTGTTAAGATAAAAAAGATTATTATTTTTGTTGAAGGTGTATCTTAAAGGAGCAA 389
Db        116 ATTTTTCTATTTTTCATTTAAATTTTTTTTTTTTTAAAAAAAAATAAAAAAAAATACTAA 57

Qy        390 ATT[illegible]AAAAACGA[illegible]A[illegible] 437
Db         56 AACTAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1

RESULT 8
US-09-960-352-6976
; Sequence 6976, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT: Jao, Nenabina
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(19298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE: 2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6976
;   LENGTH: 337
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: [illegible] LIB3058-[illegible]-Q1-K1-[illegible]
US-09-960-352-6976

  Query Match             3.3%;  Score 55;  DB 10;  Length 337;
  Best Local Similarity   53.0%;  Pred. No. 0.099;
  Matches  166;  Conservative    0;  Mismatches  140;  Indels    [illegible];  Gaps    [illegible];

Qy        147 TTTTCTTTTGGGTTTTTATTAACTGATTCAAGATTTTTATATTCTTTTTAAAGAATGGA 2[illegible]
Db         24 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 83

Qy        207 [illegible] 2[illegible]
Db         84 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTATAATTAA 14[illegible]

Qy        267 CAATCTTTTGATGTTAAGGTCAATCTTTTTGCTATGTTTTAATAAAAACAA[illegible]TTT 32[illegible]
Db        144 [illegible]
Qy        327 [illegible]
Db        204 AAAAATAA    AAAATATTAATGAAAAATGTTATTTTTTATAAATTTTAAAAAAAATG 2[illegible]

Qy        387 [illegible]AAAAA 44[illegible]
Db        2[illegible] [illegible]AAAAT[illegible]AAAAT[illegible]AT 3[illegible]

Qy        447 GATTTATATATAA 459
Db        317 TATATTTTTTTA 3[illegible]

RESULT 9
US-09-960-352-11234/c
; Sequence 11234, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT: Jao, Nenabina
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(19298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE: 2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11234
;   LENGTH: 419
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 48 LIB3058-005-Q1-K1-D8
US-09-960-352-11234

  Query Match             3.3%;  Score 55;  DB [illegible];  Length 419;
  Best Local Similarity   48.1%;  Pred. No. [illegible];
  Matches  186;  Conservative    0;  Mismatches  200;  Indels    [illegible];  Gaps    [illegible];

Qy        103 TAGATCTTTTAGTAAAAGATTTTCTTTTAGGAAGTAATGAGGGTTTTCTTTTGGGTTTTTT 1[illegible]
Db        [illegible] TATATATTTAAAAAAAGATAAAATT[illegible] 4[illegible]

Qy        163 TATTAAGTGATTCAAGATTTTTATATTCTTTTTAAAGAATGGAGTTTATTATGTAAAT 2[illegible]
Db        346 AAAAAATTTTTTTTAAAATTTTTATTTATTATTTATTTAAATTATAATTTTTTTAAAAA 28[illegible]

Qy        22[illegible] [illegible]
```

[illegible]

```
RESULT 11
US [illegible]
; Sequence 6263, Application US/[illegible]
; Patent No. [illegible]
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley [illegible]
;  APPLICANT: [illegible]
;  APPLICANT: Hyatt, [illegible]
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: [illegible] WITH LACTATION AND
;  TITLE OF INVENTION: [illegible]
;  FILE REFERENCE: [illegible]
;  CURRENT APPLICATION NUMBER: [illegible]
;  CURRENT FILING DATE: [illegible]
;  NUMBER OF SEQ ID NOS: [illegible]
; SEQ ID NO [illegible]
;   LENGTH: 4[illegible]
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: [illegible]
US [illegible]

  Query Match [illegible] Score [illegible] Length 4[illegible]
  Best Local Similarity 48.1%; Pred. No. [illegible]
  Matches [illegible] Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

[illegible]

Qy  447 [illegible]
Db  331 [illegible] 342

RESULT 12
US-[illegible]
; Sequence 7, Application US/[illegible]
; Patent No. US2002[illegible]
; GENERAL INFORMATION:
;  APPLICANT: Green, Allan
;  APPLICANT: Shah, [illegible]
;  APPLICANT: [illegible]
;  TITLE OF INVENTION: Method of Modifying the Content of [illegible]
;  FILE REFERENCE: 4[illegible]
;  CURRENT APPLICATION NUMBER: [illegible]
;  CURRENT FILING DATE: [illegible]
;  PRIOR APPLICATION NUMBER: [illegible]
```

```
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 5086
;   TYPE: DNA
;   ORGANISM: Gossypium sp.
;   FEATURE:
;   NAME/KEY:
;   LOCATION: (3785)..(5086)
;   NAME/KEY:
;   LOCATION: (3884)..(4999)
US-09-837-751-7

  Query Match             3.2%;  Score 52.6;  DB 10;  Length 5086;
  Best Local Similarity   48.9%;  Pred. No. 0.45;
  Matches  229;  Conservative    0;  Mismatches  232;  Indels    7;  Gaps    4;

Qy         13 TTTCAAAAAAGTTTTTTTATTTTCATTCTAGCAAGGGATTTGTTGCTAATTTAAGATATT 72
Db       2148 TTGAAAAAAATGAAATTTATATACTTTTTACAATTTTAGAATTATATATTTTAGAATTT 2089

Qy         73 TAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAGATTTTCTTTTAG 132
Db       2088 TA  TAAAGAAAATTATTTTTAAAACTTTTTTATAATTTTTACAAATGTTTTACATTTT 2031

Qy        133 GAAGCAATGAGGGGTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGATTTTTATATTCTT 192
Db       2030 TAATAATATTATTAATTTCAACTAATTTTTTTTCAATATTGTATTTTTGTAAATTTTTAT 1971

Qy        193 TTTTAAAGAATCCACTTTATTATTTAAATTTTTCAATACAATTTTATGATT   TTTTAA 250
Db       1970 TTTTATAGATTTTTTTATAATTTATATTTTTTTTATATTTTTTTATTTTTTAATGTGTA 1911

Qy        251 TAATGGGATTATAATGAATTTTTTAGTTAAGTTAATCTTTTTCTATGTTTTAAT 310
Db       1910 TATACATTTAATTAAACAAAATATTTTTTTTTTACATAAAGACGCCACGTGGCGCTTTGTG 1851

Qy        311 AAAAACAATGGGTTTTAACTATCTGGTAAGATAAAAAAGATTATTATTTTGTTGAAGG 370
Db       1850 ATTGGTATAAGATTATTTTTTTAATGTCTGTCAAAAAAAGGAATTTGAGAGTATTTTAG 1791

Qy        371 GTGTATCTTAAAGG-- AGCAAAGTTGATAAATTAAATTTTACTAATAAAAATAATTAAA 427
Db       1790 GTATCAAATTGAGGCAAAGAAAAGTTTAGATACCAAAGTCAAAAAATGAGTATACTTCAA 1731

Qy        428 AAACGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTACATG 475
Db       1730 GGGCCAAATTGTGAATTTAGCTTTATTTTTAATAATAGTGTTCTATG 1683

RESULT 13
US-09-960-352-13489
; Sequence 13489, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(15298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13489
;   LENGTH: 347
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 58 BoVMSI 010-01-E1-G10
US-09-960-352-13489

  Query Match             3.2%;  Score 52.6;  DB 10;  Length 347;
  Best Local Similarity   55.1%;  Pred. No. 0.29;
  Matches  103;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        380 AAAAGAATAAATTTGATAAATTAAATTTTACTAATAAAAATAATTAAAAAATGAAAATTT 439
Db         42 AAAATAATAAAAAAAAAAAAAAAAAATATAAAAAATAAAAAAAAAAAAAAAAAAAAAAAA 101

Qy        440 TATAAAAGATTTATATATAAGGAGTTGGTTTACATGCAAAAAT
Db        102 AAAAAAAAAAAATAAAAAAAAAAAAAAA

Qy        500 TTTTTTTATTATTATCCACTGTTCTTTTTGCTCAAGAGACTGAGGATTAGCAGATTTTCT 559
Db        162 ATTGATTTTTAATAAAAATTAATTATTTTTTATTAATTTTAAATTTTTTTGGGGGGTT 221

Qy        560 AAAAGGG 566
Db        222 TTCAGGG 228

RESULT 14
US-09-960-352-13784
; Sequence 13784, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(15298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13784
;   LENGTH: 397
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 59-LIB4057-004-Q1-K1-J4
US-09-960-352-13784

  Query Match             3.2%;  Score 52.6;  DB 10;  Length 397;
  Best Local Similarity   48.5%;  Pred. No. 0.27;
  Matches  177;  Conservative    0;  Mismatches  184;  Indels    4;  Gaps    1;

Qy        101 TTTAGATCTTTTAGTAAAAGATTTTCTTTTAGGAAGCAATGAGGGGTTTTCTTTTGGGTT 160
Db         20 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 79

Qy        161 TTTATTAAGTGATTCAAGATTTTTATATTCTTTTTAAAGAATCCACTTTATTATTTAAA
Db         80 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Qy        221 TTTTCAAGACAATTTTATGATTTTTTAATAATGGGATTATAATGAATTTTTTAGTT
Db        140 TTTTTTTTTTTTTTTTTTTTTTTTTTTATAAGAATTTATATATAATAAAATTTTTTTTT

Qy        281 AAGTTAATCTTTTTCTATGTTTTAATAAAAACAATGGGTTTTAACTATCTGGTAAGA 340
Db        200 TATAAAAAAATTTTTTATAATTATAAAAAAAAATAAT    TTTAATATATTAATGA

Qy        341 TAAAAAAGATTATTATTTTGTTGAAGGGTGTATCTTAAAGGAGCAAAGTTGATAAAT
Db        259 AAAGTATTTATTTACAATAATAATAGAATATATATTAGAATATAAAATAATATATAACA

Qy        401 AAATTTTACTAATAAAAATAATTAAAAAAACGAAAATTTTATAAAAGATTTATATAA
Db        319 ATATATAAAAATTTTTAATTTTATATAAAAAATTTTTTTTTTTTTTTTTTTTTTTTT

Qy        461 CAGTT 465
```

ASSOCIATED WITH LACTATION AND

[illegible]

ALIGNMENTS

RESULT [illegible]
CNS0[illegible]
LOCUS CNS0[illegible] [illegible] DNA [illegible]
DEFINITION Drosophila melanogaster genome survey sequence [illegible] BAC
BACR[illegible] of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL0[illegible]
VERSION AL0[illegible] GI:[illegible]
KEYWORDS GSS.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to [illegible])
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted ([illegible] JUN [illegible]) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr -
Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information,
please see http://www.fruitfly.org the BDGP. Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's lab (part of the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial

```
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism "Drosophila melanogaster"
                     /db_xref "taxon:7227"
                     /clone "BACR29B23"
                     /clone_lib "RPCI 98"
                     /note "end : T7"
BASE COUNT      419 a     91 c     60 g    299 t     232 others
ORIGIN

  Query Match             5.9%;  Score 98.6;  DB 17;  Length 1101;
  Best Local Similarity   37.8%;  Pred. No. 5.9e-06;
  Matches  [illegible];  Conservative  [illegible];  Mismatches  279;  Indels  [illegible];  Gaps  [illegible];

Qy          1 ATGATAATCTTTTTCAAAAAAGGTTTTTTATTTTCATTCTAGCAAGGGATTTGTTGCTA 60
Db       1068 ATAAWTAATTWWTWTATATWTAWWATATWATATATTAWTTTWTAAAATATAWWTTAATAA [illegible]

Qy         61 ATTTAAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATTTTTAGTAAAAG 120
Db       1008 ATATATWTAWTAWWATATATAWATAWAAWATATAWAWTWWAATWAATWWATAWAWTTTAAT [illegible]

Qy        121 ATTTTCTTTTAGGAAGCAATGAGGGGTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGAT 180
Db        948 AAWAATATWAWWWTWTTATTTTTTTTTTTTTTWTTTWTATWTWTTTATATWWTWWAAAAWA 889

Qy        181 TTTTATATTCTTTTTAAAGAATCGAGTTTATTATGTAAATCTTTCAAGAGAATTTTATG 240
Db        888 WWAAWTWTATTTTTTTWTTATWATTWATATTAWATTTTTWTWTATTWWTATATWATWWTA 829

Qy        241 ATTCTTTTAATAATGGTGATTATAATGAATCTTTTGATGTTAAGGTCAATCTTTTTGCTA 300
Db        828 TWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTWTATWTATAWWWTWWATWWWTAWA 769

Qy        301 TTTCTTTTAATAAAAACAATGGGTTTTAACTATCCTGGTAAGATAAAAAAGATTATTATTT 360
Db        768 AWTATATATWWATATTATWWTWATTWTTWTTTTATTWTTTAAWTTATWTATWATTWTTW 709

Qy        361 TTTTTGAAG[illegible]AAA[illegible]ATAAA[illegible]AAA[illegible]TAATA 415
Db        708 ATWTTTTTTTTTTTTTTWTATTTWAAATTATTTAATTAWWWAATWAWTTATTTAATAAW 649

Qy        416 AAAATAATTAAAAAACTAAAATTTTATAAAAGATTTATATATAAAGATTTGTTTACATG 475
Db        648 TAATTAAATTATTAAAAATTTTATWAAAAAATATTTTTTTATATTAATWAWATTTATAWA 589

Qy        476 AAAAGAAA[illegible]TAAAATTATTTTATTTTTTTTATTATTACTGTTTTTTTTTGCTCAA 532
Db        588 TTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAAATTATWTTTTTTTTTAWTAW 532

RESULT 2
[illegible]
[illegible]
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI 98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   [illegible]
VERSION     AL060786.1  GI:4949084
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information,
            please see http://www.fruitfly.org The BACR Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI 98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism "Drosophila melanogaster"
                     /db_xref "taxon:7227"
                     /clone "BACR29B23"
                     /clone_lib "RPCI 98"
                     /note "end : T7"
BASE COUNT      419 a     91 c     [illegible] g    29[illegible] t     [illegible] others
ORIGIN

  Query Match             5.7%;  Score 94.6;  DB 17;  Length 1101;
  Best Local Similarity   34.1%;  Pred. No. 2.1e-05;
  Matches  198;  Conservative  106;  Mismatches  277;  Indels  [illegible];  Gaps  [illegible];

Qy         82 AACAAAATTTGAAAGATAATTTAGATTTTTAGTAAAAGATTTTTTTTTTACTAAATAATT 141
Db        445 MMMMHAMATYTTTAKTWTMMMMMSWWAATWTWWAAAWAAAWTTATWAATWAAAAAWAAW 504

Qy        142 AGGGTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGATTTTTATATTTTTTTTAAAAA 201
Db        505 WWATTTTWWWWWWKTTTWTTTWAWWTWTAWTAAAAAAAAAAWTTTATTTTAAAWWAATAWA 564

Qy        202 ATGGAGTTTATTATGTAAATCTTTCAAGAGAATTTTATGATTTTTTTAATAATTTGATT 261
Db        565 TTAAWAATTTAWAAWWTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTTTTW 624

Qy        262 ATAATGAATCTTTTGATGTTAAGGTCAATCTTTTTGCTATGTTTTTAATAAAAACAATGG 321
Db        625 ATATAATTTTAATATTTAATTAWTTATTAATAATWWTAATTWWWTATTAAATAATTTTTW 684

Qy        322 GTTTTAACTATCCTGGTAAGATAAAAAAGATTATTATTTTTTTTTGAAGATTTATATTTAA 381
Db        685 AAATAWAAAAAAAAAAAAAAAAWAAWAAWAAWAATWATTWATTAWTTAAAAWAATTAAATWAWA 744

Qy        382 AGGAGCAAAGTTGATAAATTAAAATTTTACTAATAAAAATAATTAAAAAACTAAAATTTTA 441
Db        745 ATWAWWATAATATWWATATATAWTTWTAWWWAAAWAAAAWTTAATTAAATAWAAWWAAAW [illegible]

Qy        442 TAAAAGATTTATATATAAGGAGTTTGTTTACATGAAAAGGAAAGTAAAAGTATTTTATT 501
Db        [illegible] ATAAAAWATAWATWAAAAWAWAA[illegible]AWWA[illegible]AAAAAAAAAAAA[illegible] [illegible]

Qy        502 TTTTTTATTATTCACTGTTCTTTTTGCTCAAGAGATGATGGATTAGCAGAGGGTGTAA 561
Db        [illegible] AATWATAAWAAAAAAAWAWTTWWTWTTTTWWAWWA[illegible]AAAAWAATAAWAT[illegible] [illegible]

Qy        562 AAGGGAGAGCTGGAGAATTAGTCTAGATTTTGCCGAGCTTGCAAGAGATCAATTT 621
Db        [illegible] AAAAA[illegible]AAAAWWW[illegible]WAA[illegible]AAAAA[illegible]WWA[illegible]WAWA[illegible] [illegible]

Qy        622 AACTAGATTGATCTTACAAATTATTTGATTATTTATATTT 661
Db        [illegible] ATWTATATATWWTAWTAWATATATTTATTAAWWTATATTTT [illegible]
```

[illegible]

```
RESULT 4
BH[illegible]
LOCUS       BH[illegible]
DEFINITION  [illegible] Schistosoma mansoni [illegible]
            DNA sequence.
ACCESSION   BH[illegible]
VERSION     BH[illegible]
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; [illegible]
            [illegible]
REFERENCE   1  (bases 1 to [illegible])
  AUTHORS   Le Paslier,M.[illegible], Pierce,R.[illegible], Merlin,[illegible]
            [illegible]
  TITLE     Construction and characterization of [illegible]
            bacterial artificial chromosome library
  JOURNAL   Genomics [illegible]
  MEDLINE   [illegible]
COMMENT     Other_GSSs: [illegible]
            Contact: Pierce RJ
            INSERM [illegible]
            Institut Pasteur de Lille
            1 rue du Professeur A. Calmette, [illegible] Lille, France
            Tel: [illegible]
            Fax: [illegible]
            Email: Raymond.Pierce@pasteur-lille.fr
            GSS sequencing [illegible]
            [illegible]
            Seq primer: M13 reverse primer
            Class: BAC ends
            High quality sequence [illegible]
FEATURES             Location/Qualifiers
     source          1..[illegible]
                     /organism="Schistosoma mansoni"
                     /strain="Puerto Rican"
                     /db_xref="taxon:[illegible]"
                     /clone="[illegible]"
                     /clone_lib="SmBAC1"
                     /sex="mixed"
                     /dev_stage="[illegible]"
                     /lab_host="[illegible]"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; [illegible]
                     HindIII digested and size selected [illegible]
                     DNA was ligated into HindIII [illegible] BAC11 [illegible]
                     and used to transform [illegible] The complete library
                     contains [illegible]
                     [illegible]
                     ranges from [illegible] kb and [illegible]"
BASE COUNT  [illegible]
ORIGIN

  Query Match             [illegible]
  Best Local Similarity   [illegible]
  Matches [illegible]  Conservative [illegible]  Mismatches [illegible]  Indels [illegible]  Gaps [illegible]
```

[illegible]

```
Qy    167 AA[illegible]AAA[illegible]TTA[illegible]AAA[illegible] 226
Db    812 [illegible]TTW[illegible]AAA[illegible]TTW[illegible]AAT[illegible] 753

Qy    227 AAGA[illegible]AAT[illegible]AAT[illegible] 286
Db    752 AAAW[illegible]ANA[illegible]ATT[illegible]AAAWA [illegible]

Qy    287 [illegible]AACTA[illegible]AAGATAAA 346
Db    [illegible] AAAAAAA[illegible]AAAW[illegible]TAWA[illegible]AAAAW[illegible]

Qy    347 AAAGA[illegible]AAA[illegible]AAA[illegible] 406
Db    [illegible] TAAAAW[illegible]AAAAA[illegible]WA[illegible]AA[illegible]AAAAW [illegible]

Qy    407 [illegible]AAAAA[illegible]AAAA[illegible]AT[illegible] 466
Db    572 TA[illegible]AAATWAAAAA[illegible]AAAWT[illegible]AKAAAAAAKWA [illegible]

Qy    467 G[illegible]TA[illegible]AAAA[illegible]AAAA[illegible]AT [illegible]
Db    512 [illegible]AAWA[illegible]TT 47[illegible]
```

```
RESULT [illegible]
[illegible]
LOCUS       [illegible]    886 bp    DNA     linear   GSS 08-OCT-2001
DEFINITION  T3 end of clone 008DF11 of library SmBAC1 from strain Puerto Rican
            of Schistosoma mansoni, genomic survey sequence.
ACCESSION   AL614235
VERSION     AL614235.1  GI:16027499
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
REFERENCE   2  (bases 1 to 886)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible] 2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     [illegible]
            DNA was ligated into Hind III digested pBeloBAC 11 vector and used
            to transform E. coli DH10B. The complete library contains 23808
            clones from 4 independent sizing ligation transformations. Average
            insert size ranges from 7[illegible] 127 kb and genome coverage is 7.[illegible] fold.
FEATURES             Location/Qualifiers
     source          1..886
                     /organism="Schistosoma mansoni"
                     /strain="Puerto Rican"
                     /db_xref="taxon:6183"
                     /[illegible]
                     /[illegible]
                     /note="end : T3"
BASE COUNT      37[illegible] a     22[illegible] c     49 g     288 t     15[illegible] others
ORIGIN

  Query Match            [illegible]%;  Score 88.4;  DB 17;  Length 886;
  Best Local Similarity  4[illegible]%;  Pred. No. [illegible]23;
  Matches  177;  Conservative  [illegible];  Mismatches  187;  Indels  [illegible];  Gaps  [illegible];

Qy    [illegible]07 [illegible]TTTAG[illegible]AAAA[illegible]AAG[illegible]TTT[illegible]TTTGG[illegible]TTTTTAT[illegible] 16[illegible]
```

```
Db    872 [illegible]AAAAAAW[illegible]WAAAA[illegible]

Qy    167 AA[illegible]TTTAAA[illegible]AAAT[illegible] 226
Db    812 [illegible]TTW[illegible]AAAA[illegible]AAW[illegible]AAT[illegible]WTA[illegible]

Qy    227 AAGA[illegible] 286
Db    752 AAAW[illegible]ANA[illegible]ATT[illegible]AAAT[illegible]AAAAWA [illegible]

Qy    287 [illegible]
Db    [illegible] AAAAAAA[illegible]AAAW[illegible]AAAAWTAWA[illegible]AAAAW[illegible]

Qy    347 [illegible]
Db    [illegible] AAAAWA[illegible]

Qy    407 [illegible]AAAAA[illegible]AAAAA[illegible]AAAT[illegible] 4[illegible]
Db    [illegible]

Qy    4[illegible] [illegible]A A[illegible]AAAA[illegible]AAAA[illegible]TAT[illegible]TAT [illegible]
Db    512 [illegible]KTT[illegible]K[illegible] 4[illegible]
```

```
RESULT 6
CNS00DKY
LOCUS       CNS00DKY     928 bp    DNA     [illegible]    GSS  [illegible]
DEFINITION  Drosophila melanogaster genome survey sequence [illegible] end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL07186[illegible]
VERSION     AL07186[illegible].1  GI:494817[illegible]
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            [illegible] http://www.[illegible] The RPCI Drosophila
            melanogaster BAC library was prepared by K[illegible] and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://[illegible]/drosophila [illegible]
FEATURES             Location/Qualifiers
     source          1..928
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
                     /clone_lib="RPCI-98"
                     /note="end : [illegible]"
BASE COUNT      262 a     70 c     84 g     [illegible] t     [illegible] others
ORIGIN
```

BASE COUNT
ORIGIN

Query Match
Best Local Similarity
Matches 211; Conservative 84; Mismatches

RESULT 8
CNS01ZOM
LOCUS CNS01ZOM DNA MAY
DEFINITION Tetraodon nigroviridis genome survey sequence
sequence.
ACCESSION AL1
VERSION AL1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1 2)
AUTHORS Roest Crollius,H., Jaillon,O.,
Bernot,A., Dasilva,C., Wincker,P.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome-wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1 2)

```
AUTHORS     Roest Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
TITLE       Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 1029)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (12 APR 2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/tetraodon
FEATURES             Location/Qualifiers
     source          1..1029
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="22G11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : CGAG22CB6E06SP1 end :
                     PUC orf"
BASE COUNT      297 a    124 c     66 g    457 t     85 others
ORIGIN

  Query Match             5.2%;  Score 89.2;  DB 17;  Length 1029;
  Best Local Similarity   43.5%;  Pred. No. 0.88845;
  Matches  272;  Conservative  128;  Mismatches  173;  Indels   ;  Gaps   ;

Qy          9 CTTTTTTCAAAAAAGGTTTTTATTTTATTTTAGGAAGGATTTTTTGTAATTTAAGA 68
Db        402 CTGTTTTTTATTTCTTATATATTTATTTTTTTTTAATTTTTTTATTTTTTTTTTWTTT 461

Qy         69 TTTTAAGTATTAATAAAATTTAAATATAATTTATATTTTTATTAAAAAA....T 128
Db        462 TTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTT 521

Qy        129 TAGGAAGCAATGAGGGTTTTTTTTTGGGTTTTTATTAAGTGATTCAAGATTTTTATAT 188
Db        522 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTATATATATATTTATAT 581

Qy        189 TTTTTTTAAAGAATGAGTTTATTATGTAAATTTTCAAGAGAATTTTATGATTCTTT 248
Db        582 TTTTATATATAAATAAAAATAAAAAAAATAATTATATATTAATTTTWTTTTTTTTTTTTT 641

Qy        249 AATAATGGTGATTATAATGAATTTTTGATGTTAAGGTCAATCTTTTTGCTATGTCTTTA 308
Db        642 ATTAAATTWTTTWTATAATAATTTTAAAAATTTTTTTAAAAATTTTAAGTATTAAAAAAA 701

Qy        309 ATAAAAACAATGGGTTTAACTATTTGGTAAGATAAAAAAGATTATTATTCTTGTTGAA 368
Db        702 AAAWAAAAWAAAATATTTTTATAATTTTTTWTTWAWTWTWWATATATTAWTTTTWTTTA 761

Qy        369 GGGTATTTAAATATAAATTTATAAATTAAATTTTACTAATAAAAATAATTAAAA 428
Db        762 AAATATWTTTTTATAAATATATAAAWWAATATTTATTTTWATATWATAWATTTTTTTWTW 821

Qy        429 AACTAAAATTTTATAAAAGATTTATATATAAGGATTTGGTTTACATGAAAAGGAAAGTTA 488
Db        822 AATAATAAAAATAATAAWATTTTWTKTTTWATTATTAATATNTKTATATATWTWTATWTA 881

Qy        489 AAATTATTTTATTTTTTTTATTATTACTGTTCTTTTTGCTCAAGAGACTGAT 541
Db        882 ATATATTTTTATTTTTTTTTTTAAWAWATAWAWATWATTAATAAAAAAATTTWT 934

RESULT 9
CNS00FUHZ/c
LOCUS       CNS00FUHZ              996 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC
            BACR41B21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071063
VERSION     AL071063.1  GI:4951105
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     BAC end sequencing was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org. The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..996
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR41B21"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      383 a    164 c     81 g    171 t    197 others
ORIGIN

  Query Match             5.1%;  Score 87.6;  DB 17;  Length 996;
  Best Local Similarity   36.6%;  Pred. No. 0.00056;
  Matches  210;  Conservative   99;  Mismatches  270;  Indels    ;  Gaps    ;

Qy          6 AATCTTTTTTCAAAAAAGGTTTTTATTTTCATTCTAGGAAGGATTTTTTGTAATTTA 65
Db        995 AAAWTTTTTTWATAAAAAAAAATWTTTTWWTWATTWMAAAAWTWAWAWATWAMATTWWTA 936

Qy         66 AGATATTTAAGAGATGAAGAAAATTTGAAAGATAATTTAGATTTTTATAAAAGATTT 125
Db        935 ATTTKTTTKTTTTTTTTTTTTTTTTKAAAAAAAAAAAAAAAWWWATTTTTWTWAAW 876

Qy        126 TTTTAGGAAGCAATGAGGGTTTTTTTTTGGGTTTTTATTAAGTGATTCAAGATTTTTA 185
Db        875 TTTAARTTWRATTTWAWTAWWTAWWTTTTTTTTTTTTTTTTTTTTAATWTTTTWTTTTTTTW 816

Qy        186 TATTTTTTTTAAAGAATGAGTTTATTATGTAAATTTTTAAGAAATTTATGATTTT 245
Db        815 ATTTWATWWWTMAAWAWWTTAWTTWWTTWATATAAWY     AAATWTTTTTTTTTTTTT 761

Qy        246 TTAATAATGGTGATTATAATGAATTTTTGATGTAAGGTCAATTTTTTTTTATTTTT 305
Db        760 TTTKATWAATTAWTTTATTTTWTATTTTTTWWTATTWWTTTTTTATTWTTTTTTTTTTT 701

Qy        306 TTAATAAAAACAATGGGTTTAACTATTTGGTAAGATAAAAAAGATTATTATTTTTTT 365
Db        700 TTWWTWATTTTTTWWTATTTATWTTAWTGTATTKAWTTAAAGWTTTWTTTTTTKTWK 641

Qy        366 GAAGGGTATTTAAATATAAATTTATAAATTAAATTTTACTAATAAAAATAATTAAAA 425
Db        640 TTWATAATWAATTTWTAWTTAKWAWTTWKAATTTTWTAAAWTWTTTTTTTWAAATATKKA 581

Qy        426 AAAAACTAAAAT....ATAAAAGATTTATATATAAGGATTTGGTTTACATGAAAAGGAAAT 485
Db        580 TTWAWWTTTWATKTTTTTTTATTTATTATTTWTKAYTTTTTTTTKTTTTTTTTTTTTATT 521
```

[illegible]

```
Db    [illegible] AAAAAAAAAAAAAAAAAAAAA [illegible]
Qy    [illegible] AAAAAAAAAAA [illegible]
Db    [illegible] AATAAATTAWAATTTWTTTATWTAAWAWWTTW [illegible] AATAAAAWAAAAAA [illegible]
Qy    [illegible] AAAAAAAAAAAAAAAA [illegible]
Db    [illegible] AAAAATTTTATAAATAKAT [illegible]
Qy    [illegible]
Db    [illegible] TTWTAAATWATTTKTNAT [illegible] AAAAWAWAAT [illegible]
Qy    [illegible] ATTAAAAAAA [illegible] TAAAAAAATTTATAAAAAAT [illegible]
Db    [illegible] TWATTTTATAAWAATT [illegible] AAAAAAAAAAA [illegible]
Qy    [illegible] AAATTAAAAATTATTTTATTT TTTTATTATTTATT [illegible]
Db    1052 AWAAWTAAAAWWAAWWWAATTTAAATAATWTAAAAT [illegible]
```

RESULT [illegible]
[illegible]

```
LOCUS       [illegible]     DNA     [illegible]
DEFINITION  Drosophila melanogaster genome survey sequence [illegible] BAC
            BACR[illegible]N[illegible] of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   [illegible]
VERSION     [illegible]
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases [illegible])
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted [illegible] JUN [illegible] Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information,
            please see http://www.fruitfly.org. The RPCI-98 Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y; cn bw sp, the same strain used for the BDGP
            P1 and BAC libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila.htm
FEATURES             Location/Qualifiers
     source          [illegible]
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR[illegible]"
                     /clone_lib="RPCI-98"
                     /note="end : [illegible]"
BASE COUNT  [illegible]
ORIGIN

  Query Match [illegible]
  Best Local Similarity [illegible]
  Matches [illegible] Conservative [illegible] Mismatches [illegible] Indels [illegible] Gaps [illegible]

Qy    [illegible] AATTTAATATATTTAATATATAATAAAAATTTAAATATAATTTATA [illegible] TATTAAAA [illegible]
```

```
Db      45  [illegible]
Qy     120  [illegible]TTATTAA[illegible]                          179
Db     510  AAAAATATAA[illegible]TTAA                          569
Qy     180  TTTTATATTCTTTTTTAAAGAATGGAGTTTATTATGTAAATCTTTCAAGAGAATTTTAT 239
Db     570  TTTTAAWAWAAA[illegible]AAWWAAAAAAAAAWTTTA 629
Qy     240  [illegible]TAA[illegible]                          299
Db     630  AAWAWTW[illegible]WAAAAAAAWATAWAAWTTW[illegible]WAAATWA 689
Qy     300  [illegible]AAGATAAAAAAGATTATTATT                      359
Db     690  AA[illegible]AAAAWTTATTAWTTTTAAAAWATWAWTW 749
Qy     360  [illegible]TAAATTTTACTAATAAAAA 419
Db     750  WTAAAT[illegible]TAAWATTAWAAAAT[illegible]TAATTTTTTTTAATTAAAA 808
Qy     420  TAATTAAAAAACGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTACATGAAAA 479
Db     809  AAAAAAAAAAAAAAWAAATAAAAAAATAAAAAAAAAAAWAAAWWTTTTTTTAAATTATTT 868
Qy     480  [illegible]CTTTT[illegible]CAAGAGA[illegible] 539
Db     869  WTTATTTAAAAATTTTAATAATSATTTAAWTAT[illegible]AAGTWTAAAAAWTGKGAATG 928
Qy     540  ATGGATTAGCAGAGGGTTCTAAAAGGGCAGAGCCTGGAGAATTACTCTTAGA 591
Db     929  KKKF [illegible]TKAATKKTAAA 980
```

RESULT 12
CNS06QXV/c
LOCUS   CNS06QXV   [illegible] bp   DNA   linear   GSS [illegible] 2001
DEFINITION 17 end of clone AW0AA009H09 of library AW0AA from strain CLIB 89 of
            Yarrowia lipolytica, genomic survey sequence.
ACCESSION   AL411257
VERSION   AL411257.1 GI:[illegible]
KEYWORDS   GSS.
SOURCE   Yarrowia lipolytica.
  ORGANISM Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1 (bases 1 to 759)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., [illegible] Balaguer-[illegible],
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE   Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 1-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2 (bases 1 to 759)
  AUTHORS   Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE   Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
            lipolytica
  JOURNAL   FEBS Lett. 487 (1), 95-100 (2000)
  MEDLINE   20584727
   PUBMED   11152892
REFERENCE   3 (bases 1 to 759)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT   This GSS is part of a random genomic sequencing program of 13 [illegible]
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and [illegible] the sequence of
            the other extremity of this insert.
FEATURES   Location/Qualifiers
     source   1..759
                     /organism="Yarrowia lipolytica"
                     /strain="CLIB 89"
                     /db_xref="taxon:4952"
                     /clone="AW0AA009H09"
                     /clone_lib="AW0AA"
                     /note="end : 17"
BASE COUNT   [illegible]
ORIGIN

  Query Match   [illegible]; Score 84.4; DB [illegible]; Length [illegible];
  Best Local Similarity   41.5%; Pred. No. [illegible];
  Matches 194; Conservative   64; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Qy   5 TAAT[illegible]CAAAAAA[illegible]TAA[illegible] 64
Db   480 TTWT[illegible]TWTCKCKKTKKTKWT[illegible]TKWTTKA [illegible]
Qy   65 AAGATATTTAAGAGATGAACAAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAGATTT 124
Db   420 [illegible] [illegible]
Qy   125 [illegible] 184
Db   360 [illegible] [illegible]
Qy   185 [illegible] 244
Db   300 TTTTTTTTTPTAATTTTTTTGTTTTTTTTPTTTTTTTTTTTTTKTTTTTTKKKTKTTTTT 241
Qy   245 [illegible]
Db   240 [illegible]
Qy   305 [illegible]
Db   180 [illegible]
Qy   365 [illegible]
Db   120 [illegible]
Qy   424 TAAAAAACGAAAATTTTATAAAAGATTTATATATAAGGAGTTGGTTTA 471
Db   64 TTAAAAATTAAAAATAAATTAAAAATTAAAAATTAAAAAGAAATTTTA 17
```

RESULT 13
CNS03BD
LOCUS       CNS03BD                 1101 bp    DNA     linear   GSS [illegible]
DEFINITION  Drosophila melanogaster genome survey sequence [illegible] end of BAC #
            [illegible] of [illegible] library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064[illegible]
VERSION     AL064091.1  GI:4941847
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
```

This BAC end sequence was carried out as part of a ... Berkeley ... Genome Project (BGP).

... Drosophila ...

... further information ...

... Department of ... Institute in Buffalo, ...

... used for the BGP's ...

... of the library ... entire library ...

Matches ... Gaps

```
LOCUS       CNS0410K                         RNA
DEFINITION  Tetraodon nigroviridis ...
            library ...
            sequence.
ACCESSION   AL28?149
VERSION     AL28?149.1
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; ...
            Actinopterygii; Neopterygii; ...
            Acanthopterygii; ...
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to ...)
  AUTHORS   ...
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate ...
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to ...)
  AUTHORS   ...
            Weissenbach,J.
  TITLE     Characterization and repeat analysis ...
            freshwater pufferfish Tetraodon ...
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to ...)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (...)
COMMENT     This sequence is ...
            scale ... end sequencing project ...
            genome. For more information, please ...
            http://www.genoscope.cns.fr/...
FEATURES             Location/Qualifiers
     source          1..
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:..."
                     /clone="..."
                     /clone_lib="..."
                     /note="Genoscope ..."
BASE COUNT
ORIGIN

  Query Match
  Best Local Similarity
  Matches

Qy          1 ATGATAA...AAAAAA...
Db        ... ATTTA...
Qy         61 ATTTAAGATATTTAAGAGATGAACAAAA...
Db        ... NTTTTATWNAATAWATAAW...
Qy        121 ATTT...
Db        ... TWTATT...
Qy        181 TTTTATATT...AAA...
Db        ... WTTTWTT...
Qy        241 ATT...AATAATT...
Db        ... TWTATWTWTT...
Qy        ... TTTTTAATAAAAAA...
Db        ... AATT...
```

```
Qy        361 [illegible]TAAATTTTACTAATAAAAAT 420

Db        531 [illegible]AAAAAAWWWTWTTATTTTTTTATTTWTWTTTTWAWAATWNTTTAAWTTTTATTAAT 472

Qy        421 AATTAAAAAAAGAAAATTTTATAAAAGATTTATATATAAGAGTTGGTTTACATGAAAAG 480

Db        471 TATAAAAAAAAWAAWWWTAAWKNTANTAAAAAAAAANWAADKKTGTNKTAAATGNTTRTT 412

Qy        481 GAAAGCTAAAAGTATTTTATTTTTT 505

Db        411 AAAWDNBNAAAAAMNNNNNNYGATT 387

RESULT 17
[illegible]
LOCUS       [illegible]              1101 bp    DNA     [illegible]
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BAC[illegible] of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL[illegible]
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02 JUN 1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp. the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http:[illegible]/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR08K13"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      201 a     64 c     131 g    202 t     503 others
ORIGIN

  Query Match             5.0%;  Score 83.6;  DB 17;  Length 1101;
  Best Local Similarity   14.3%;  Pred. No. 0.0011;
  Matches   79;  Conservative  205;  Mismatches  150;  Indels   [illegible];  Gaps   [illegible];

Qy         25 TTTTTTATTTTCATTCTAGTAAGGGATTTGTTTCTAATTTAAGATATTTAAGAGATGAAC 84

Db       1041 WWWWWWATWDTWWTKWWWWATAAKTTAWTWWKTAWKAKWAGKKKGAKKDKGAATTADT 982

Qy         85 AAAATTTGAAAGATAATTTAGATCTTTTAGTAAAAGATTTTCTTTTAGGAAGCAATGAGC 144

Db        981 AGPRPPPPRRRRRRRRR[illegible]RKRKAARAARWATKWWDKWKWKRKWKWRAKDRK 922

Qy        145 GGTTTTCTTTTGGGTTTTTATTAAGTGATTCAAGATTTTTATATTTCTTTTTAAAGAATG 204

Db        921 [illegible] 862

Qy        205 GAGTTTATTATGTAAATTTTTAAGAGAATTTTATGATTTTTTTAATAAT[illegible]GATTATA 264

Db        861 GWADADWWT---WDAAADWWAD[illegible]WTAWAKWWT[illegible] 8[illegible]

Qy        265 ATGAATTTTTTGATTTAAGTTTAATTTTTTGTTATTTTTTTAATAAAAACAATG[illegible] 324

Db        804 KRPDRKRADKRDAADRDAAIWTIWTTTTTRTTDWKWKTTWTRWAATRTWTRTTT 745

Qy        325 TTAACTATCCTGGTAAGATAAAAAAGATTATTATTCTTGTTGAAGGGGTATCTTAAAGG 384

Db        744 TRDRAGTACKKWRDWKRKWKRTTPWTDAD[illegible] 6[illegible]

Qy        385 ACTAAAGTTTATAAATTAAATTTTACTAATAAAAATAATTAAAAAACGAAAATTTTATAA 444

Db        684 [illegible]AAWWTTTTTTTTWT[illegible]AAWTWKA 6[illegible]

Qy        445 GAGATTATATATAATTATTTTTTTTATATGAAAATTAAATTTAAAATTTTTTATTTT 504

Db        624 WWAKWWWKTTAKWKTWAA[illegible] 5[illegible]

Qy        505 TTTATTAT 512

Db        564 TTTATWTT 557

Search completed: November 4, 2002, 22:54:[illegible]
Job time : [illegible]74.33 secs
```